

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOTEHNOLOOGIA ÕPPETOOL

Geneetiline varieeruvus mikroRNA geenides

Magistritöö (30 EAP)

Mirjam Tamme

Juhendaja: Tarmo Annilo (PhD)

TARTU 2017

INFOLEHT

Geneetiline varieeruvus mikroRNA geenides

MikroRNAd (miRNAd) on olulised kõigis põhilistes rakulistes protsessides, sealhulgas rakkude jagunemisel, diferentseerumisel ja apoptoosis. Seondudes sihtmärk-mRNAle moduleerivad nad mRNA tööd. Geneetiliste variantide poolt põhjustatud häired selles regulatsiooniprotsessis võivad viia erinevate haiguste tekkeni.

Ühenukleotiidsed muutused miRNA järjestustes omavad inimese kui terviku funktsioneerimises olulist rolli ning seega on oluline omada võimalikult palju informatsiooni nende asukohtade ning mõju kohta miRNA funktsioneerimisele.

Antud töö raames on vaadeldud ühenukleotiidsed variatsioone (SNVsid) mikroRNA geenides Eesti Geenivaramu andmestiku põhjal. Kindlaks on tehtud miRNA geenides asuvad SNVd, nende asetsemine mikroRNA funktsionaalsetes osades, samuti SNVde tihedus vastavates piirkondades. Veel on töös käsitletud mikroRNAde haigusseosisust ning kahe mikroRNA põhjal on ka vaadeldud SNVde konserveeritust läbi liikide

Märksõnad: mikroRNA; Ühenukleotiidsed variatsioonid

CERCS teaduseriala: B220 Geneetika, tsütogeneetika

Genetic variation in microRNA genes

MicroRNAs (miRNAs) are important factors in all main cell processes, including cell division, cell differentiation and apoptosis. Through binding to their target mRNA they may inhibit it's work, which may lead to development of different diseases. Single nucleotide variations (SNVs) in these small molecules might have a strong impact on the functioning of human organism, which makes it critical to obtain more information about their locations and impact.

In the following work we have observed SNVs in miRNA genes based on the Estonian Genome Center dataset. We have identified SNVs located in miRNA genes, as well as the density of these variations in different functional parts of a microRNA. In addition, we identified relationship between SNV density in a microRNA gene and it's association to a disease. Finally, based on two microRNAs, the relationship of evolutionary conservation and SNVs through different species were described.

Keywords: microRNA; single Nucleotide Variation

CERCS research project: B220 Genetics, cytogenetics

SISUKORD

| | |
|---|-----------|
| INFOLEHT | 2 |
| SISUKORD..... | 3 |
| KASUTATUD LÜHENDID JA TERMINID | 5 |
| SISSEJUHATUS..... | 6 |
| 1.KIRJANDUSE ÜLEVAADE | 7 |
| 1.1 MikroRNAd | 7 |
| 1.1.1.MiRNAd biogenees..... | 7 |
| 1.2 MiRNAd ja haigused | 9 |
| 1.2.1 MiRNAd ja vähkkasvajad..... | 10 |
| 1.3MiRNAd geneetiline varieeruvus | 10 |
| 1.3.1MiRNAd ja ühenukleotiidsed muutused..... | 11 |
| 1.3.2 Ühenukleotiidsed muutused miRNAdes vähimarkeritena..... | 11 |
| 1.3.3MiRNAd ja koopiaarvu muutused | 12 |
| 1.3.4 Geneetilise variantide seosed fenotüüpidega | 12 |
| 1.4 miRNAd klastrid | 13 |
| 1.5Meetodid SNVde uurimiseks miRNAdes..... | 13 |
| 1.5.1MiRNAd SNVde andmebaasid..... | 14 |
| 1.6 miRNAd evolutsioon | 15 |
| 1.7Suuremad populatsiooniuuringud | 16 |
| 2.EKSPERIMENTAALNE OSA..... | 17 |
| 2.1 TÖÖ EESMÄRK | 17 |
| 2.2 MATERJAL JA METOODIKA | 17 |
| 2.2.1 Töövoo üldkirjeldus | 17 |
| 2.2.2 Valimi kirjeldus | 17 |
| 2.2.3 MiRNAd genoomsed koordinaadid | 18 |
| 2.2.4 MiRNA geenides olevate variantide tuvastamine..... | 19 |
| 2.2.5 Funktsionaalsed miRNA osad..... | 19 |
| 2.2.6 Variantide määramine miRNA struktuuris | 19 |
| 2.2.7 SNVde tiheduse definitsioon | 20 |
| 2.2.8 miRNA ja haiguste vahelised seosed..... | 20 |
| 2.3 TULEMUSED JA ARUTELU..... | 21 |

| | |
|---|-----------|
| 2.3.1 MiRNA geenide variatsioon Eesti Geenivaramu andmestiku põhjal | 21 |
| 2.3.2 Haigusseoselised SNVd miRNAdes | 27 |
| 2.3.3 miRNA alusjärjestuses paiknevad SNVd evolutsioonilise konserveerumise taustal.. | 31 |
| KOKKUVÕTE | 34 |
| SUMMARY | 35 |
| TÄNUSÕNAD | 36 |
| 3.KASUTATUD KIRJANDUS | 37 |
| KASUTATUD VEEBIAADRESSID | 45 |

KASUTATUD LÜHENDID JA TERMINID

AF – *Allele frequency*, alleelisagedus

Alusjärjestus – *The seed sequence*, miRNA seondumist määrav järjestus

3' UTR – 3' *untranslated region*, 3' mittekodeeriv regioon

HC_miRNA – *High Confidence microRNA*, kõrge usaldusväärsusega miRNA

EGV – Eesti Geenivaramu

eQTL - *Expression quantitative trait loci*, ekspressiooni kvantitatiivse tunnuse lookused

GWAS - *Genome-wide association study*, ülegenoomne assotsiatsiooniuring

miRNA – mikroRNA

mRNA – *messengerRNA*, informatsiooni-RNA

pre-miRNA – prekursor miRNA

pri-miRNA – primaarne miRNA

SNV – *Single nucleotide variation*, ühe nukleotiidne muutus

SISSEJUHATUS

MikroRNAd (miRNAd) on väikesed mittekodeerivad RNAd, millel on oluline funktsioon geeniregulatsioonis. Nad omavad tähtsust kõigis põhilistes rakulistes protsessides, sealhulgas rakkude jagunemisel, diferentseerumisel ja apoptoosis. MikroRNAd reguleerivad kodeerivate geenide ekspressiooni post-transkriptsioonilisel tasemel ning seondudes oma sihtmärk-mRNAGA võivad takistada selle tööd, mille tagajärjel võivad alguse saada mitmed erinevad haigused.

MiRNAd avastati eelmise sajandi lõpus ning alates sellest ajast on nende mõju uuritud suures mahus. MiRBase'i andmebaas, mis haldab endas miRNA gene erinevatest liikidest, laieneb pidevalt. Siiski leiame me pidevalt aina uusi aspekte miRNA funktsioonide ning mõju kohta organismi, kui terviku, funktsioneerimisele. Selle põhjuseks võib pidada asjaolu, et miRNA seondumise pea ainsaks aluseks on komplementaarsus miRNA alusjärjestuse ning mRNA seondumissaidi vahel, mis annab võimaluse mitmetel mRNAdel olla ühe miRNA sihtmärgiks või samuti ühel mRNAl olla mitme miRNA sihtmärgiks. Üheks vähem uuritud aspektiks miRNade puhul on variatsioonid nende geenides ning nende mõju miRNA edasisele funktsioneerimisele. Siiski ühenukleotiidsed muutused miRNA geenides võivad omada edasisele protsessile suurt mõju, avaldades mõju seondumisele sihtmärk geenile, mistõttu see ei saa toimida enam nii, nagu mitte muutunud alleeli puhul.

Antud magistritöö eesmärk on kindlaks teha Eesti Geenivaramu andmestiku põhjal ühenukleotiidsed muutused, mis mõjutavad mikroRNAsid ning nende seondumist sihtmärgile. Seoses sellega antakse ülevaade nende asukohtadest miRNA funktsionaalsetes osades ning hinnatakse nende mõju miRNA funktsiooni muutumisele.

1. KIRJANDUSE ÜLEVAADE

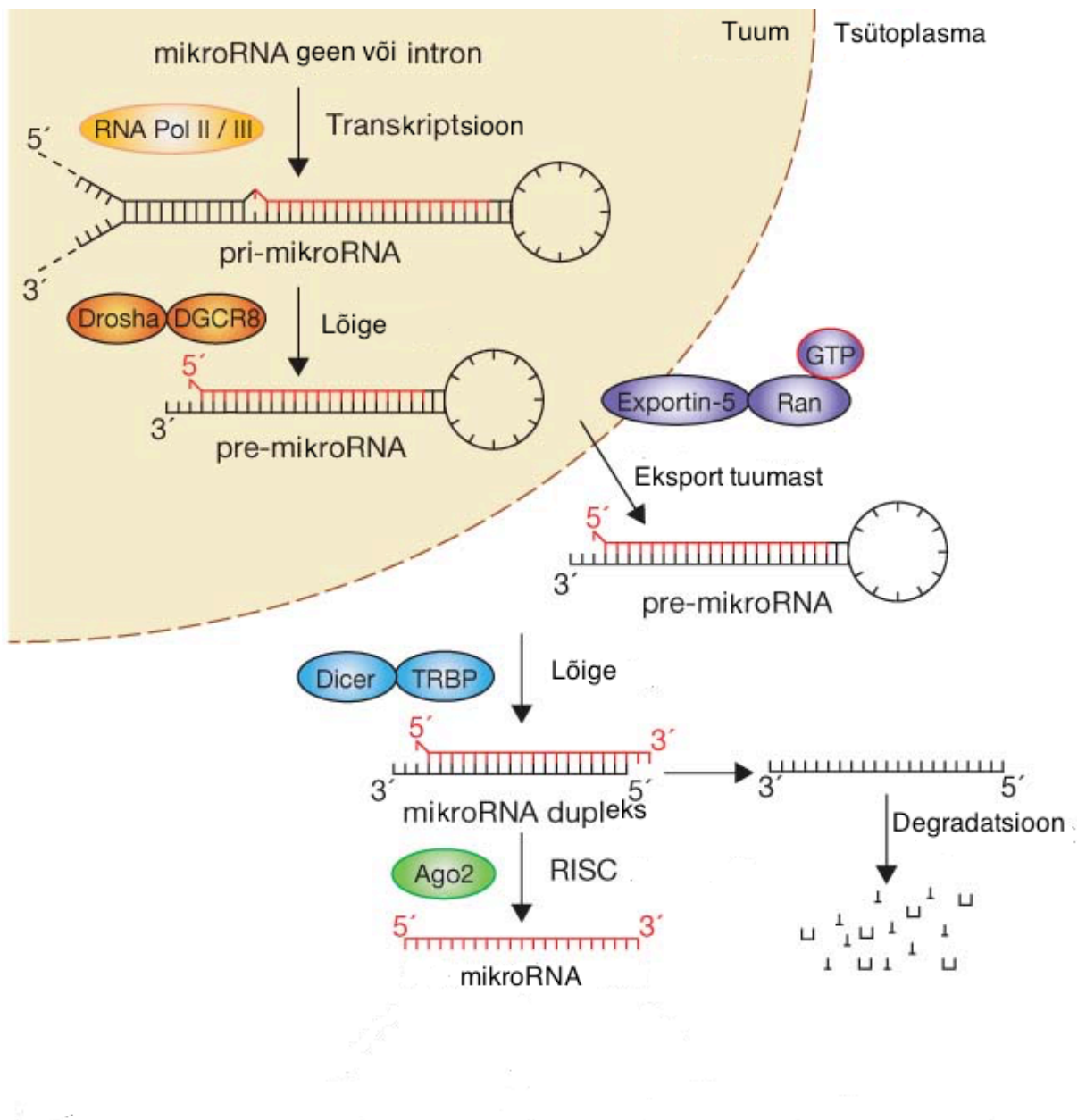
1.1 MikroRNAd

MikroRNAd (miRNAd) moodustavad suure perekonna keskeltläbi 21 nukleotiidi pikkustest mitte-kodeerivatest RNAdest. Nad on üheks oluliseks geenide avaldumise regulaatoriks loomades ja taimedes. Imetajates kontrollivad miRNAd kolmandikku valku kodeerivate geenide aktiivsusest (Rajewsky, 2006). Samuti osalevad nad peaaegu kõigis rakulistes protsessides. Paardudes mRNAga vahendavad miRNAd translatsioonilist repressiooni või mRNA degradatsiooni (Filipowicz *et al.*, 2008).

MiRNade aktiivsust täheldati esmakordselt *C.elegans*'i geenides, kui märgati, et varbussi varajast arengut kontrollivates geenides on olulised tegurid kaks väikest regulaatorset RNAd, lin-4 ja let-7 (Lee *et al.*, 1993). Let-7 homolooge täheldati aastal 2000 ka kõrgemates organismides (Reinhart *et al.*, 2000), andes alust arvata, et let-7 ja ka teised väikesed RNAd on paljudes liikides olulisel kohal (Pasquinelli *et al.*, 2000). Varsti peale seda leiti lin-4 ja let-7 RNAsid erinevates organismides: ussides, kärbestes ja imetajates. Neid väikeseid RNAsid hakati nimetama mikroRNAdeks (Lagos-Quintana *et al.*, 2002). Tänapäevaks on selge, et miRNAsid leidub lisaks loomadele nii taimedes, rohevetikates kui ka viirustes (Griffiths-Jones *et al.*, 2008).

1.1.1. MiRNade biogenees

MiRNade transkriptsioon võib toimuda nii intronite kodeerivast alast kui ka teistelt geenidelt, nagu näiteks teistelt mitte kodeerivatelt RNAdelt või ka kodeerivate geenide eksonitelt. Enim toimub miRNade transkriptsioon siiski nende endi geenidelt.



Joonis 1: miRNAde biogenees. Tuumas transkribeeritakse RNA polümeraasi II või III abil pri- miRNA. Transkript lõigatakse ensüümkompleksi Drosha-DGCR8 abil pre-miRNAks ning viiakse eksportiin-5-Ran-GTP kompleksi abil tuumast tsütoplasmasse. Dicer-TRBP kompleks lõikab pre-miRNA-d nii, et tekib miRNA dupleks. Dupleksi funktsionaalne osa (punane) ühineb Ago2 valgu abil RISC kompleksiga, mis tunneb sihtmärkgeenide 3'-UTR järjestustes ära komplementaarsed saidid ja valgu ekspressioon surutakse maha. Märklau-mRNA ning miRNA vahelise komplementaarsuse ulatusest sõltub, millise mehhanismi kaudu toimub geeniekspressiooni kontroll (kohandatud Winter *et al.*, 2009 järgi).

Mõned miRNAd, mis asetsevad intronites jagavad enda reguleerivaid elemente ja primaarseid transkripte pre-mRNA peremees geenidega. Ülejäänud miRNA geenid arvatakse olevat transkribeeritud nende enda promootoritest (Lee, 2002).

RNA polümeraasideks, mis võivad pri-miRNA transkriptsiooni läbi viia on polümeraas II ja polümeraas III (Johnson *et al.*, 2003).

Peale transkriptsiooni lõikab mikroprotsessor kompleks pri-miRNA katki. Mikroprotsessor kompleks koosneb RNase III Drosha ensüümi ja tema RNAGA seonduvast partnerist, DGCR8. Tulemuseks on juuksenõela struktuur, mille pikkus on keskel läbi 65 nukleotiidi. Seda struktuuri nimetatakse pre-miRNAks. Pre-miRNA struktuuri tunneb ära Exportin-5 (XPO5) ensüüm, mis transpordib molekuli tsütoplasmasse, kus toimub töötlemine ATP-sõltuva RNAase III-e Dicer ensüümi poolt (Yi *et al.*, 2003). Diceri töö tulemuseks on lühike 19-24 aluspaarine miRNA dupleks. Selle dupleksi juhtivahel moodustab RNA indutseeritud vaigistamise kompleksi (RISC), mis sisaldab Argonaut valke põhikomponentidena. RISC kompleksis olev miRNA toimib kui giid, tundmaks ära rakulist mRNA-d, et indutseerida nende degradatsiooni või inhibeerida translatiooni (Joonis 1).

1.2 MiRNAd ja haigused

Alates inimese täisgenoomi kindlaks tegemisest on teadus aina rohkem keskendumas mittekodeerivate RNA-de uurimisele ja nende rollile inimese genoomis. Nende väikeste RNA-de roll inimeste haigustes teeb neist olulised molekulid, mida kliiniliste uuringute valguses on vaja mõista. MiRNAd on olulisel kohal väga paljudes arenguprotsessides, nagu näiteks metabolism, rakkude proliferatsioon, apoptoos, arengu ajastus ja neuraalsete rakkude saatus. Teised reguleerivad rollid on neuraalne geeniekspressioon, aju morfogenees, lihaste diferentseerumine ja tüvirakkude jagunemine (Mattick ja Makunin, 2005).

Nüüdseks on tehtud kindlaks miRNA-de oluline roll protsessides, mis on seotud eukarüootide arenguga. Tänu nende potentsiaalile kontrollida rakkude kasvu ja diferentseerumist on nad head kandidaadid vähi teraapias (Bartel, 2004).

MiRNA-de vähesust või ülehulka on seostatatud mitmete kliiniliselt oluliste haigustega, nagu näiteks müokardi infarkt ja autoimmuunhaigused. Ühenukleotiidsed mutatsioonid miRNA-s või tema sihtmärgis on mehhanismiks, mis mõjutavad miRNA-de efekti rakus (Szymanski *et al.*, 2005).

1.2.1 MiRNAd ja vähkkasvajad

Inimese miRNA geenid asuvad tihti genoomsetes piirkondades, mida seostatakse vähiga. Nendes miRNAdes esineb suurel hulgal erinevaid mutatsioone, mille mõju on erinevatele vähitüüpidele erinev. Kõige sagedamini esineb ühenukleotiidseid mutatsioone (transitsioone ja transversioone), harvemini deletsioone ja insertioone. Nende mutatsioonide esinemissagedus miRNAdes varieerub sõltuvalt vähitüübist. Sagedamini esineb mutatsioone miRNAdes, mis on seotud melanoomi, endometriaalse-, rinna- ja maovähiga ning harvemini hepatotsellulaarse kartsinoomi, medulloblastoomi ja kilpnäärmevähi puhul (Tuna *et al.*, 2016), mille põhjal saab väita, et esmalt mainitud vähitüüpide puhul võib miRNAdel olla juhtiv roll vähi levikus ning arengus.

Kindlaks on tehtud, et vähiga seotud miRNAd on rohkem konserveerunud ning nende SNVde sagedus on madalam võrreldes mitte vähiga seotud miRNAdega. Samuti on näidatud, et vähiga seotud miRNAd asetsevad rohkem klastrites, kui mitte vähiga seotud. Antud aspektid võivad viidata sellele, et vähkkasvajate tekkeks on vajalik ka mitmete erinevate miRNade koostöö. SNVd esinevad neis siiski harvemalt, kuid asetsevad positsioonides, mis on mRNAle seondumiseks kriitilise tähtsusega, olles seeläbi oluliseks faktoriks vähkkasvajate tekkes. Seega, on miRNAd üheks võimalikuks viisiks vähi levikut diagnoosida (Zhou *et al.*, 2015).

1.3 MiRNade geneetiline varieeruvus

Paljud uuringud on sidunud miRNAsid mitmete erinevate bioloogiliste protsessidega, sealhulgas raku kasv, proliferatsioon, diferentseerumine ning areng (Kloosterman ja Plasterk, 2006). Selle loogilise järelalusena seostatakse miRNA deregulatsiooni erinevate haigustega nagu vähk, aga ka psühhiaatriliste- ja neuroloogiliste haigustega (Lu M *et al.*, 2008). Nagu ka teistel funktsionaalselt olulistel geneetilistel elementidel oodatakse ka miRNAdes pigem väiksemat geneetilist varieeruvust ning vähem mutatsioone. Suurel enamusel mutatsioonidel võib olla negatiivne efekt nende funktsionaalsusele või biogeneesile (Sun G *et al.*, 2009). Seega võivad haigus-seoselised SNVd miRNA tasemeid mõjutada väga erinevatel viisidel. Nii näiteks võivad mutatsioonid miRNA biogeneesi geenides mõjutada miRNA protsessimist, SNVd pre-miRNAdes miRNA valmimist ning SNVd miRNA promooter alas juba valmis miRNA ekspressiooni (Hrdlickova *et al.*, 2014).

1.3.1 MiRNAd ja ühenukleotiidsed muutused

Erinevad *in silico* analüüsid ühenukleotiidsete muutuste kohta inimese miRNA geenides on näidanud, et miRNA geenidel on madalam SNVde tihedus, kui nendega külgnevatel regioonidel (Saunders *et al.*, 2007; Gong *et al.*, 2012; Han ja Zheng, 2013).

MiRNA geeni siseselt on valmis järjestusel (*mature miRNA*) madalam SNVde tihedus kui pre-miRNAdel. Sealjuures on seondumist määraval järjestusel ehk alusjärjestusel (*the seed sequence*) madalaim SNVde tihedus, näidates nende funktsionaalset olulisust. Samuti on näidatud negatiivse korrelatsiooni esinemist SNVde arvu ja haiguste vahel, mida seostatakse miRNAGA (Han ja Zheng, 2013). See omakorda tõestab geneetiliste variantide olulisust miRNA geenides seoses nende funktsiooniga ning seotust inimese haigustega.

SNVd on indiviidide erisuses väga olulise tähendusega, mängides rolli nii paljude fenotüüpide, omaduste ning haiguste tekkes (Shastri, 2009). Erinevad muutused miRNAdes endi geenides mõjutavad nende biogeneesi, funktsionaalse ahela valikut ning sihtmärgi valikut. Praeguseks hetkeks on kindlaks tehtud väga mitmete miRNA sihtmärkide ning miRNA geenide SNVde seos erinevate haigustega (Ryan *et al.*, 2010). On näiteks kindlaks tehtud, et SNV pre-miR-146a-s vähendab valmis miRNA ekspressiooni ning soodustab papillaarse kilpnäärme kartsinoomi teket (Jazdzewski, *et al.*, 2008). Samuti on näidatud, kuidas mutatsioon inimese miR-96-s põhjustab progressiivset kuulmis-kadu (Mencia, *et al.*, 2009).

1.3.2 Ühenukleotiidsed muutused miRNAdes vähimarkeritena

Genoomi-ülesed uuringud on teinud kindlaks mitmeid geneetilisi markereid genoomi mittekodeerivates osades. Kaks märkimisväärset näidet on vähkkasvajate tekkega seostatud piirkonnad kaheksanda kromosoomi pikas õlas (8q24.2) ning 11. kromosoomi pikas õlas (11q13.3) (Tumbull *et al.*, 2010). Nendes piirkondades on kindlaks tehtud 37 miRNA geeni. Nendes miRNAdes on uuritud näiteks ühte SNVd miR-1206-es (rs2114358) ja kahte SNVd miR-612s (rs550894, rs12803915), mille puhul on täheldatud seotust vähiga (Kim *et al.*, 2012). Antud töödes on miR-612-e ekspressiooni täheldatud rakutüübi spetsiifiliselt; ekspressioon suureneb SNVde esinemise puhul eesnäärme vähi rakuliinides, väheneb aga käärsoole vähi rakuliinides ning ei oma mingit efekti rinnavähi rakuliinidele. SNV (rs2114358) miR-1206 –es mõjutab samuti miRNA ekspressiooni määra, kuid seda mitte rakuliini spetsiifiliselt (Kim *et al.*, 2012).

1.3.3 MiRNAd ja koopiaarvu muutused

Koopiaarvu variatsioonid (CNVd) on genoomse piirkonna koopiate arvu varieerumine indiviidide vahel (Iafrate *et al.*, 2004).

CNVde ning miRNAde koevolutsiooni ning interaktsioone peetakse katseks saavutada tasakaal doosi-tundlikes geenides, samal ajal suurendades mitmekesisust mitte-tundlikes geenides, panustades seeläbi liikide evolutsiooni. *In silico* ülegenoomsed analüüsid on näidanud, et nii miRNA geenide enda kui ka nende seondumissaitide arv on CNV piirkondades kõrgem, kui mitte CNV piirkondades. (Felekis *et al.*, 2011).

Kõike seda arvesse võttes võime oletada, et miRNAd on tekkinud evolutsiooni käigus mehhanismina tõstmaks tolerantsi genoomi plastilisuse suhtes (Dweep *et al.*, 2013).

1.3.4 Geneetilise variantide seosed fenotüüpidega

Mitmed uuringud on näidanud korrelatsiooni polümorfismide vahel pri-miRNAdes (Hogg ja Harries, 2014) ning vastava hulga miRNAdes. Näiteks vaadates geneetilisi variante inimese pre-miRNAdes, mis on seotud rinnavähiga on esile tõstetud kaks haruldast varianti pre-miRNAdes, miR-30c ja miR-17 (Shen *et al.*, 2009). Sellega kaasnevad muutused nende sekundaarstruktuuris, mis omakorda viib ebakorrapärase miRNA ekspressioonini.

Viise, kuidas järjestuse variatsioon võib miRNAde aktiivsust mõjutada on mitmeid. Nende hulka kuuluvad:

- 1) variatsioon järjestustes, mis määravad miRNA biogeneesi;
- 2) variatsioonid pre- või pri-miRNA järjestuses;
- 3) variatsioon miRNA promootor alas;
- 4) variatsioon järjestuses, mis on vajalik seostumiseks mRNAle.

Nii näiteks variatsioonid *RAN* (rs14035), *DROSHA* (rs10719), *XPO5* (rs11544382) ja *DICER* (rs3742330) geenides, mis on ühed põhilised miRNA biogeneesi eest vastutavad geenid, on seostatud idiopaatilise korduva raseduse katkemisega (Jung *et al.*, 2014) ning geeni *XPO5* variant rs11544382 on seostatud esmase munasarja puudulikkusega (Rah *et al.*, 2013). Samuti on variatsioone miRNA protsessimise masinavärgis seostatud vähiga. Seda põhjusel, et kaks SNV'd *GEMIN4* geenis (rs2740348 ning rs7813) võivad olla seotud suurenenud riskiga neeru kartsinoomi tekkeks (Horikawa *et al.*, 2008) ning ühenukleotiidne mutatsioon rs11077, mis asub *XPO5* geenis ja on seostatud hepatotsellulaarse kartsinoomiga (Liu *et al.*, 2014).

1.4 miRNAde klastrid

MiRNAde kogumid, mis asetsevad genoomis üksteise lähedal nimetatakse miRNAde klastriteks. Klastrisse arvestatakse miRNAd, mille kaugus teineteisest jääb 10kb vahemikku (Shalgi *et al.*, 2007). Selle raames on kindlaks tehtud, et kaugused miRNA paaride vahel genoomis järgivad bimodaalset jaotust ehk kõige sagedamini esinevaid väärtusi on mitu. Vahe kahe maksimumi vahel on umbes 10kb, millest saabki oletada, et 10kb võib olla mõistlik suurus miRNA klastrite defineerimiseks.

MiRNAde soodumust klastreid moodustada täheldati esmakordselt väikeste mitte-kodeerivate RNAde suuremahulises uuringus (Lagos-Quintana *et al.*, 2003).

Praeguseks on tehtud kindlaks, et miRNAde klastrid on laialt levinud just loomade genoomides (Olena ja Patton, 2010). Sellega seoses on näidatud, et üle 30% inimese, hiire, roti ja kana miRNA geenidest asuvad klastrites (Megraw *et al.*, 2007).

Samuti on täheldatud miRNA klastrite konserveerumist erinevates liikides (Altuvia *et al.*, 2005). Antud asjaolu annab kinnitust, et need klastrid on omandanud spetsiaalsed regulatoorsed funktsioonid bioloogilistes protsessides. Sarnaselt on ka ekspressiooniuuringud näidanud, et klasterdunud miRNAd ekspresseeruvad tihti koos (Baskerville ja Bartel, 2005). Seega on tõenäoline, et eksisteerib regulatoorne koordineerimine klasterdunud miRNAde vahel.

1.5 Meetodid SNVde uurimiseks miRNAdes

Meetodeid miRNAde uurimiseks on erinevaid ning olenevad, nagu iga uuringu puhul, esitatud küsimusest. Üheks mõjusaimaks viisiks SNV-de funktsionaalsuse kindlaks tegemisel on leida eQTL efekt (Hrdlickova *et al.*, 2014).

Hapmap'i konsortsiumi SNVde uuringus (*The International HapMap Consortium*, 2003) tuvastati 32 3'UTRi SNVd, mis potentsiaalselt mõjutavad miRNA seondumist (Gamazon *et al.*, 2012). Kuna miRNAdele on palju sihtmärke, siis on andmete analüüsiks vaja meetodeid, mis suudavad neid sihtmärke eristada. MiRNAde valdkonnas kuuluvad tõenduspõhiste viiside alla RNA ristsidumine RISC kompleksile, nende komplekside immunosadestamine hõivates Argonaut valgu (enim kasutatakse antikehi, mis toimivad Argonauti vastu või kasutatakse rakuliine, mis ekspresseerivad märgitud Argonaut valke), samuti sekveneeritakse mRNA sihtmärke kompleksis. Selliste meetodite näideteks on AGO2 HITS-Clip (Chi *et al.*, 2009), PAR-CLIP (Hafner *et al.*, 2010) ning CLIP-seq (Zisoulis *et al.*, 2010). Nende meetodite kitsaskohaks on, et nad ei anna spetsiifilist informatsiooni miRNA seondumise kohta kindlale sihtmärgile. Probleemi lahendamiseks on välja töötatud meetod, CLASH (Helwak *et al.*,

2013), mille puhul on tegu hübriidide ristseondumise, ligatsiooni ning sekveneerimisega. Selle meetodi korral toimub esmalt RISC komplekside ristseondumine, eraldatud Argonaudi immunosadestamisest ning seejärel ligeeritakse RNA neis kompleksides, mille tulemusel formeeruvad miRNA-sihtmärk mRNA hübriidid, mida saab sekveneerida. Tänu CLASHile on olemas ka kõrge lahutusvõimega meetod, millega identifitseerida sihtmärke, kuhu miRNA seondub ning annab ettekujutuse sellest, millised fraktsioonid RISC kompleksis on hõivatud spetsiifiliste miRNAde poolt.

1.5.1 MiRNAde SNVde andmebaasid

Selleks, et koondada erinevaid andmeid SNVde kohta miRNAde või nende seondumiskohtades on loodud mitmeid veebipõhiseid andmebaase ja vahendeid. Näiteks on olemas dbSMR (Hariharan *et al.*, 2009), mille puhul on tegu andmebaasiga, kus on inimese miRNA seondumiskohad ning SNVd, mis nendes seondumiskohtades leiduvad. Andmebaasist saab informatsiooni küsida kasutades kas geeni nime, miRNA nime, polümorfismi või transkripti ID koodi. Andmebaasist oli võimalik vaadata näiteks reguleerivate SNVde rolli 3'UTR alal ning populatsiooni spetsiifiliselt moduleerida miRNA sihtmärke.

Veel üheks andmebaasiks on miRNASNP andmebaas (Gong *et al.*, 2015), kus on välja toodud SNVd pre-miRNAde nii inimeses kui ka teistes liikides ning sihtmärgi spetsiifilisuse teke või kadu miRNA alusjärjestuses või 3'UTR regioonis sihtmärk mRNAde.

Veel üheks olemasolevaks andmebaasiks on miRvar (Bhartiya *et al.*, 2011), mille puhul on tegu kasutajate endi poolt hallatava ressursiga, käsitlemaks genoomset variatsiooni miRNA lookustes. mirVar pakub kasutaja-sõbralikku lähenemist uurimaks variante erinevates miRNAde, sisaldades informatsiooni selle variandi positsiooni ning alleeli muutuse suhtes vastavas miRNA. Andmebaasist saab samuti otsida miRNAid kasutades nende nime, haigus-seoselisust, SNV asukohta juuksenõela struktuuril ja kromosoomi.

Veel üheks andmebaasiks on mirVAS (Cammaerts *et al.*, 2015), mille puhul on tegu programmiga, mis ennustab geneetiliste variantide mõju miRNAdele. MirVAS automatiseerib ennustamise, annoteerides SNVasukoha miRNA juuksenõela funktsionaalsele regioonile ning annoteerides kõik struktuursed muutused miRNA, mis antud variandiga kaasnevad.

Paljud neist keskenduvad SNVdele just miRNAde seondumiskohtades, kuid on ka andmebaase, mis keskenduvad muutustele miRNA enda geenides ning sellele kuidas see mõjutab nende sihtmärgi valikut ja biogeneesi.

1.6 miRNAde evolutsioon

MiRNAid leidub enamike eukarüootide genoomides, alustades pruunvetikatest kuni kõrgemate loomadeni välja. Siiski meetodid, kuidas miRNAid funktsioneerivad ning kuidas neid protsessitakse, viitavad sellele, et tekkeviis on loomadel ning taimedel olnud sõltumatu (Cuperus *et al.*, 2011). Seega näiteks taimedes toimub kogu miRNA biogeneesi protsess tuumas ning Drosha ja DGCR8 homologe taimedes ei leia.

MiRNAid on kasulikud fülogeneetilised markerid, sest aja jooksul nad väga kiirelt evolutsioneerunud ei ole (Wheeler *et al.*, 2009). Regulaatorse mehhanismina pärinevad miRNAid RNA interferentsi masinavärgist, mis oli algselt kasutusel võitluses eksogeense geneetilise materjali vastu, nagu näiteks viirused (Pashkovski ja Ryazansky, 2013). MiRNAid võisid tekkida morfoloogiliste innovatsioonide kaudu, muutes geenide ekspressiooni spetsiifilisemaks. Need täiendused on lubanud aga tekkida aina keerukamatel organismidel ja seega on ka miRNAidel oma roll elu keerukuse tekkes. Samuti seostatakse järske morfoloogilisi innovatsioone kõrge miRNAde akumulatsiooniga (Heimberg *et al.*, 2008).

Uued miRNAid tekivad erinevatel viisidel. Nii näiteks võivad miRNAid pärineda juuksenõela struktuuride tekkest DNA mitte-kodeerivates alades (intronid või intergeensed regioonid), aga ka juba olemas olevate miRNAde duplikatsiooni ja modifikatsiooni teel (Nozawa *et al.*, 2010). Evolutsiooni kiirus (näiteks nukleotiidsed asendused) evolutsiooniliselt nooremates miRNAdes on võrreldavad teiste mitte-kodeerivate DNA elementidega, viidates neutraalsele evolutsioonile. Siiski vanematel miRNAidel on muutused tekkinud harvemini (tihti vähem kui üks asendus saja miljoni aasta kohta) (Peterson *et al.*, 2009). Evolutsiooniline surve toimib erinevatele miRNA regioonidele eritugevusega, kusjuures piirkonnad, mis on elutähtsad funktsioneerimiseks on enam konserveerunud (Warthmann *et al.*, 2008).

SNVde osakaal pre-miRNAdes on tunduvalt madalam kui seondumatel regioonidel ning see väheneb miRNA evolutsioonilise vanuse tõusuga. Nii pre-miRNAde kui ka miRNAde puhul on SNVde tihedus suhteliselt suurem uuemate miRNA geenide puhul, võrreldes nendega, mis pärinevad varasemast ajast (Zhu *et al.*, 2012). Järjestuse konserveeritus on tugev miRNAidel, mis pärinevad ajast enne primaatide-näriliste lahknemist, andes võimaluse oletada tugevat puhastavat valikut nii pre-miRNAdele kui ka valmis miRNAdele vanade miRNA geenide puhul. Samal ajal evolutsiooniliselt nooremate miRNAde puhul on järjestuse asenduste arv suur, mõlema nii funktsionaalsete järjestuste kui ka nende lähedaste regioonide osas, olles rohkem kui kolm korda kõrgemad kui vanemate miRNAde puhul (Zhu *et al.*, 2012).

1.7 Suuremad populatsiooniuringud

MiRNA SNV-de kindlaks tegemiseks on läbiviidud mitmeid suuremahulisi populatsiooni uuringuid. Ühe sellise põhjal on koostatud varem mainitud MirSNP andmebaas. Selles andmebaasis on kindlaks tehtud 414 510 SNVd, mis võivad mõjutada miRNA-mRNA seondumist (Liu *et al.*, 2012). Kasutades mirSNP andmebaasi kolme aju eQTL andmestikule on kindlaks tehtud neli SNVd (rs3087822, rs13042, rs1058381 ja rs1058398), mis võivad mõjutada miRNA seondumist ning seeläbi muuta nende sihtmärk geenide ekspressiooni ajus. Samuti on skisofreenia GWAS tuvastanud seitse miRNA SNVd (Liu *et al.*, 2012).

2. EKSPERIMENTAALNE OSA

2.1 TÖÖ EESMÄRK

Antud töö eesmärk on leida Eesti Geenivaramu geenidonorite andmestikust ühenukleotiidsete muutuste asukohad miRNA geenides ning hinnata nende mõju ja haigusseoselisust.

2.2 MATERJAL JA METOODIKA

2.2.1 Töövoo üldkirjeldus

Antud töös on varieeruvust miRNA geenides hinnatud erinevaid aspekte vaadeldes. Allpool on toodud ülevaade kasutatatud meetoditest.

Kasutusel on SNVde informatsioon Eesti Geenivaramu andmestikust (valimist rohkem kirjas allpool). Pre-miRNA ning miRNA koordinaadid saadi mirBASE'i andmebaasist ning nendes leiduvad SNVd filtreeriti välja kasutades UCSC andmebaasi tabel-otsingumootorit. Seejärel defineeriti funktsionaalsed pre-miRNA osad, et teha kindlaks ka varieerumise erinevus erinevates pre-miRNA osades. Kõigiks nendeks analüüsideks kasutati UCSC tabel-otsingumootorit ning filtreerimist selle alusel.

Ühe ülesandena vaadeldi töö käigus antud SNVde evolutsioonilist konserveerumist läbi selgroogsete, seda kahe pre-miRNA näitel, mille alusjärjestuses asus SNV. Selleks võeti mirBASE'ist kõik olemasolevad antud pre-miRNA järjestused ning annoteeriti inimese järjestuse alusel.

Kõigi nende sammudega loodetakse antud tööga anda esmane ülevaade varieeruvusest miRNA ning pre-miRNA geenides Eesti populatsioonis.

2.2.2 Valimi kirjeldus

Antud uurimistöö käigus kasutati kogu genoomi järjestusega Eesti Geenivaramu doonorite andmeid (N=2244). Nendest 1000 indiviidi olid andnud nõusoleku osaleda teistkordses proovide andmises (keskmine aeg proovide andmise vahel on neli ja pool aastat). Teine osa indiviididest (N=1300) valiti eesmärgiga katta võimalikult hästi Eesti populatsiooni geneetiline mitmekesisus, kasutades selleks geenidoonori enda poolt teatatud sünnikoha infot. Kogu genoomi sekveneerimise metoodika, kvaliteedikontroll ning variantide tuvastamine on detailselt kirjeldatud Guo et al. artiklis (Guo *et al.*, 2017). Pärast erinevatel etappidel

läbiviidud kvaliteedikontrolli jäi lõplikuks analüüsitavate indiviidide arvuks 2244. Kogu genoomi sekveneerimise protsessi on kirjeldatud lühidalt allpool.

Sekveneerimine viidi läbi Broad Instituudis kasutades Illumina HiSeqX aparatuuri. Iga positsioon genoomis kaardistati keskmiselt 30-kordse kattuvusega. Saadud järjestused joondati inimese referentsgenoomile (hg19) kasutades Burrows-Wheeler Aligner'it (BWA) (Li ja Durbin, 2009). Geneetilised variandid tuvastati programmiga HaplotypeCaller paketist GATK v3 (Van der Auwera *et al.*, 2013), (DePristo *et al.*, 2011), (McKenna *et al.*, 2010).

Variantide kvaliteedikontrollis rakendatud parameetrid on täpsemalt kirjeldatud Guo, *et al* artiklis (Guo *et al.*, 2017).

Uuringute läbiviimiseks on geenidoonorid andnud informeeritud nõusoleku ning olemas on Tartu Ülikooli inimuuringute eetika komitee kooskõlastus.

2.2.3 MiRNAde genoomsed koordinaadid

MiRNA informatsioon saadi miRBase andmebaasist (21 versioon). Sealne informatsioon sisaldas miRNA nime ja asukohta genoomis. Andmestikus on 1877 inimese pre-miRNA, millest 295 on kõrge usaldusväärsusega ning 2810 miRNA, millest 617 kuuluvad kõrge usaldusväärsusega andmestikku.

Selleks, et tegu oleks kõrge usaldusväärsusega miRNAga peab miRBase 21 versioonis järjestus olema kaardistatud vähemalt 10 korral mõlema õla suunal ning kokku peab olema seda järjestust kaardistatud vähemalt 100 korda. Kõrge usaldusväärsusega andmeid on vaja, kuna uute väikeste RNAde avastamise hulk kasvab hüppeliselt, tulenevalt täiustunud sekveneerimisvahenditest. Seega satuvad miRBase valimisse ka miRNA, mille ekspressioonitase on väga madal. Nad asuvad väga spetsiifilistes kudedes või ajutiselt mingis rakutsükli faasis. Seetõttu ongi vaja kõrge usaldusväärsusega valimit, mille põhjal saame kindlalt väita, et tegu on reaalse miRNAdega, mitte fragmentide või teiste transkriptidega.

Analüüsid on läbi viidud nii kogu valimi kui ka kõrge usaldusväärsusega miRNAdega.

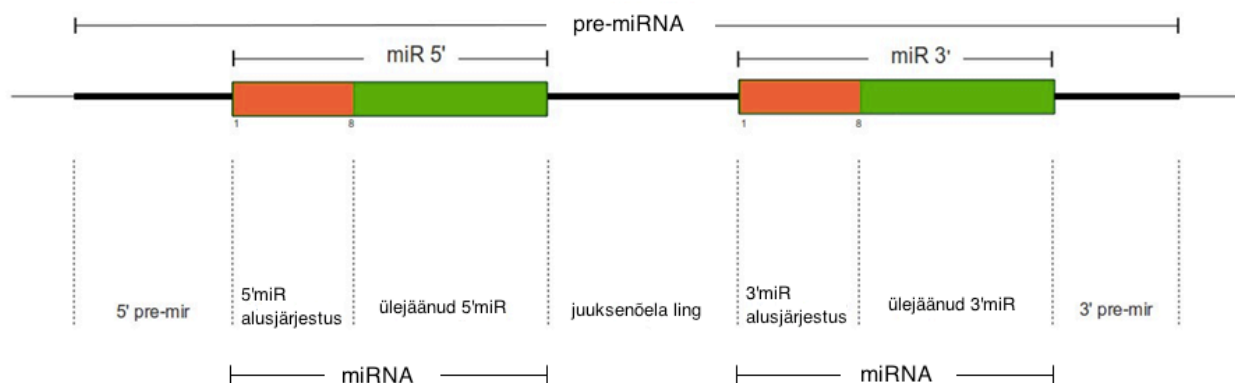
MiRNAde koordinaadid on konverteeritud UCSC andmebaasi kasutades genoomi versioonist GRCH38 versiooni GRCH37 (hg19). Seda just põhjusel kuna hg19 puhul on tegu kõige täielikumalt anoteeritud versiooniga.

2.2.4 MiRNA geenides olevate variantide tuvastamine

Ühenukleotiidsed muutused Eesti Geenivaramu täisgenoomide andmestikus tuvastati miRNA genoomsete koordinaatide põhjal. Variantide väljavõtmine teostati kasutades Mart Kalsi poolt koostatud käsujada.

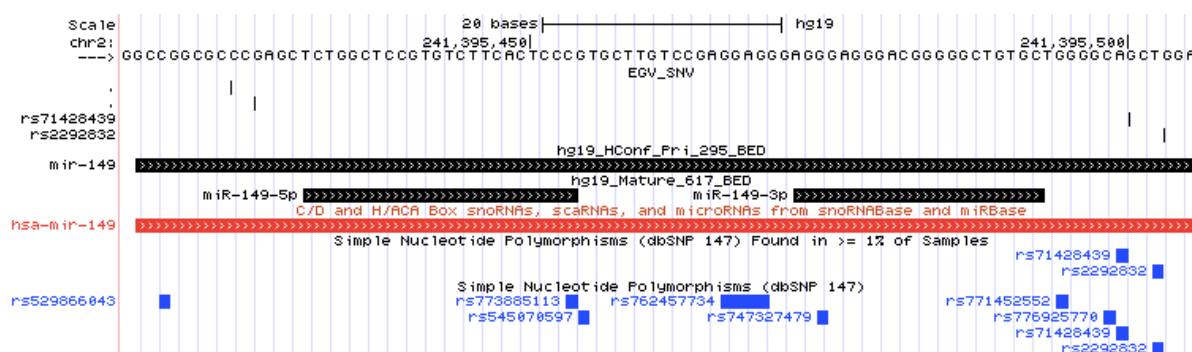
2.2.5 Funktsionaalsed miRNA osad

MiRNA funktsionaalsed regioonid määrati vastavalt joonisele 2 (joonis 2). Alusjärjestus koosneb miRNA nukleotiididest 2-8 (alustades miRNA 5' otsast).



Joonis 2. Pre-miRNA skemaatiline struktuur. Valmis miRNA koosneb 5' miR alusjärjestusest ning ülejäänud 5' miR struktuurist. Teine valmis miRNA samadest 3' miR komponentidest. Kahe miRNA vahele pre-miRNA struktuuris jääb juuksenõela struktuur. (Carbonell *et al.*, 2012).

2.2.6 Variantide määramine miRNA struktuuris



Joonis 3. Visuaalne kujutis ühest miRNA geenist (mir-149), kasutades UCSC genoomi brauserit, kus lisaks kasutaja sisestatud EGV SNVdele on välja toodud ka dbSNP andmebaasis olemasolevad SNVd. Joonisel on välja toodud hg19 referents genoomijärjestus; pre-miRNA mir-149 asukoht; miRNAde mir-149-3p ja mir-149-5p asukohad; EGV SNVde

asukohad genoomil; USCS andmebaasis olev miRNA asukoht; dbSNP andmebaasi sisestatud SNVd.

Selleks, et teha kindlaks variandi esinemine miRNA geenis sisestati UCSC genoomi brauserisse (Kent *et al.*, 2002) EGV SNVde genoomsed koordinaadid ning samuti miRNAde koordinaadid. Antud tegevuse tulemusena saadi visuaalne kujutis SNVde esinemisest miRNA geenides. Võrdluste tegemiseks kasutati sama andmebaasi tabelitöötlus vahendit (Karolchik *et al.*, 2004).

2.2.7 SNVde tiheduse definitsioon

SNVde tiheduse arvutamiseks järjestuses on kasutatud järgnevat valemit

$$SNV\ tihedus = \frac{N \times 1000}{L}$$

Kus N tähistab SNVde arvu järjestuses ning L on järjestuse (miRNA) pikkus, aluspaarides. SNVde tiheduse visualiseerimiseks kasutati R-tarkvara.

2.2.8 miRNA ja haiguste vahelised seosed

MiRNAde ja haiguste vahelised seosed laeti alla Inimese miRNA ja haiguste andmebaasist (Human MiRNA & Disease Database, HMDD), kus säilitatakse kirjandusest leitud ja inimese poolt üle kontrollitud seoseid miRNAde ja haiguste vahel (Lu *et al.*, 2008). Seisuga 23.05.2017 leidis seal 10 368 sisestust, mis sisaldasid 572 miRNA geeni ning 378 haigust 3511 kirjanduslikust allikast.

2.2.9 Statistilised analüüsid

Andmetabeleid korrastati programmi Microsoft Office Excel 2010 abil ja üldine valimit iseloomustav analüüs viidi läbi programmiga R (versioon 3.2.0) ning Microsoft Office Excel 2010-ga. Karpdiagrammide joonistamisel kasutati programmi R (versioon 3.2.0).

2.2.10 Järjestuste joondamine

Joondamiseks kasutati programmi T-Coffee ning kõik joondamised on tehtud inimese (hsa, *Homo Sapiens*) järjestuse annotatsioonil.

2.3 TULEMUSED JA ARUTELU

2.3.1 MiRNA geenide variatsioon Eesti Geenivaramu andmestiku põhjal

Eesti Geenivaramu täisgenoomide andmestikust (N=2244) leiti 1735 SNVd pre-miRNAs.

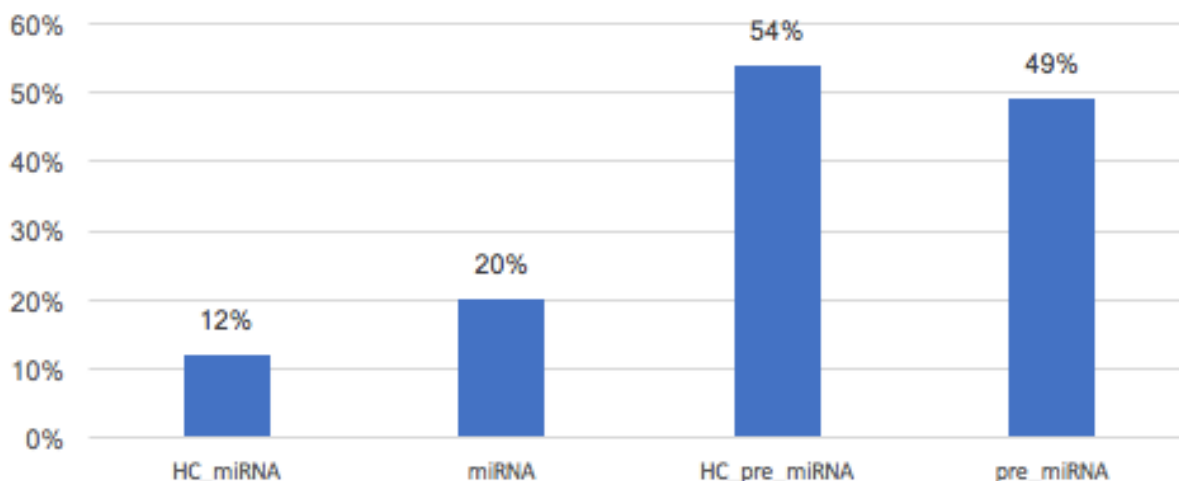
Sagedasteks variantideks on võetud alleelid, mille alleelisagedus (AF) on suurem kui 1% ning haruldasteks variantideks alleelid, mille AF on väiksem kui 1%. Eesti Geenivaramu valimis on 403 (23,2%) sagedast varianti (AF >1%) ning 1332 (76,8%) haruldast varianti (AF<1%).

Tabel 1 võtab kokku SNVde arvu ning nende paiknemise kas prekursor või küpse miRNA struktuuris.

Tabel 1. Kokkuvõtte leitud variantide positsioonidest. Valimi suurus, sekveneeritud indiviidide arv (geenodoonorite arv); SNVde arv, leitud SNVde arv valimis; dbSNV, SNVd valimist, mida on juba varem kirjeldatud; SNVd pre-miRNAs, kui palju SNVsid asusid miRNA prekursor piirkonnas (kõrge usaldusväärsusega miRNAd/mitte-kõrge usaldusväärsusega miRNAd); SNVd küpses miRNAs, kui palju SNVsid asusid küpse miRNA struktuuris; haigusseoselisus, SNVde hulk, mis on seotud mõne teadaoleva haigusega.

| miRNA | Valimi suurus | SNVde arv | dbSNV | SNVd pre-miRNAs | SNVd küpses miRNAs | Haigus-seoselisus |
|-------|---------------|-----------|-------|-----------------|--------------------|-------------------|
| EGV | | | | | | |
| valim | 2244 | 1735 | 1288 | 228/1730 | 97/655 | 754 |

Edaspidi on vaadeldud SNVde esinemist miRNAdes ja pre-miRNAdes. Samuti on eraldi esile toodud kõrge usaldusväärsusega andmestikud.

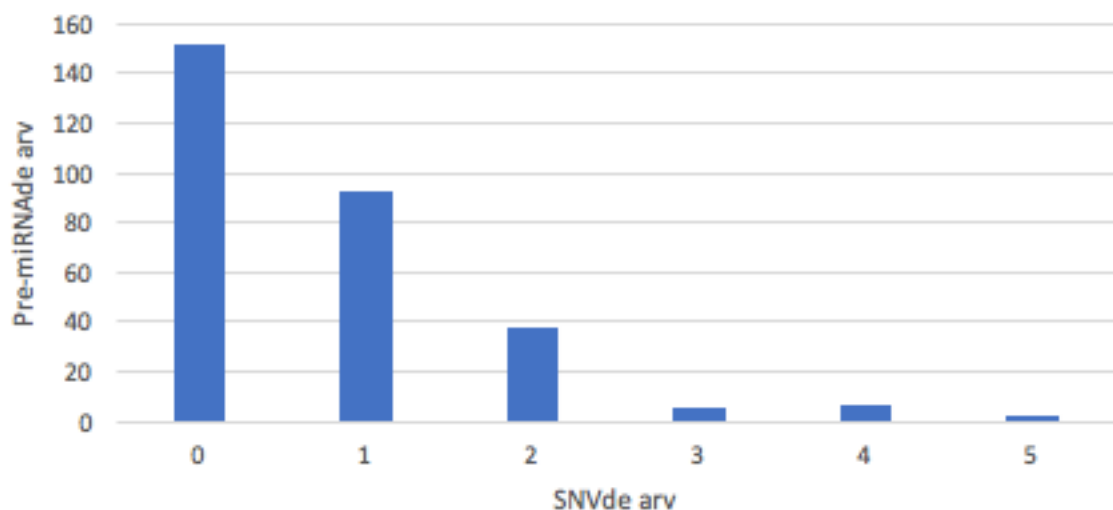


Joonis 4. SNVsid sisaldavad miRNAd ning pre-miRNAd. X-teljel on esitatud nii kõrgeusaldusväärsusega (HC) miRNAd, kui ka pre-miRNAd ning mitte-kõrge usaldusväärsusega miRNAd ning pre-miRNAd. Y-teljel on esitatud miRNAde osakaal protsentuaalselt, milles esines SNV.

Joonisel 4 on toodud välja miRNAd, milles leidis SNV EGV andmestikus. Välja on toodud protsentuaalselt nende miRNAde ning pre-miRNAde osakaal, mis sisaldasid SNVd Eesti Geenivaramu andmestikust.

HC_miRNAde puhul on tegu kõrge usaldusväärsusega andmetikuga, kus 617 miRNAs on SNV 86-s, mis esindab 12% antud miRNAdest. MiRNA andmestikus on toodud kõik hetkel mirBASE-is välja toodud miRNAd, mida on 2800, nendest sai SNV kindlaks teha 549-s. See kujutab endast 20% kõigist miRNAdest. Pre-miRNAde puhul sai kindlaks teha 1877 miRNAst SNV 10151-s miRNAs, mis esindab 49% kõigist miRNAdest. Vaadates vaid kõrge usaldusväärsusega pre-miRNAsid oli võimalik leida 295-st pre-miRNAst 144 SNVga miRNAst.

Järgnevalt vaadeldi ka ühes miRNAs leitud SNVde arvu, mille järgi saame ka pildi nende molekulide variatsioonide suurusest. Ühes pre-miRNAs olevate võimalike SNVde arv varieerus kõrge-usaldusväärsusega pre-miRNAde andmestikus nullist viieni.

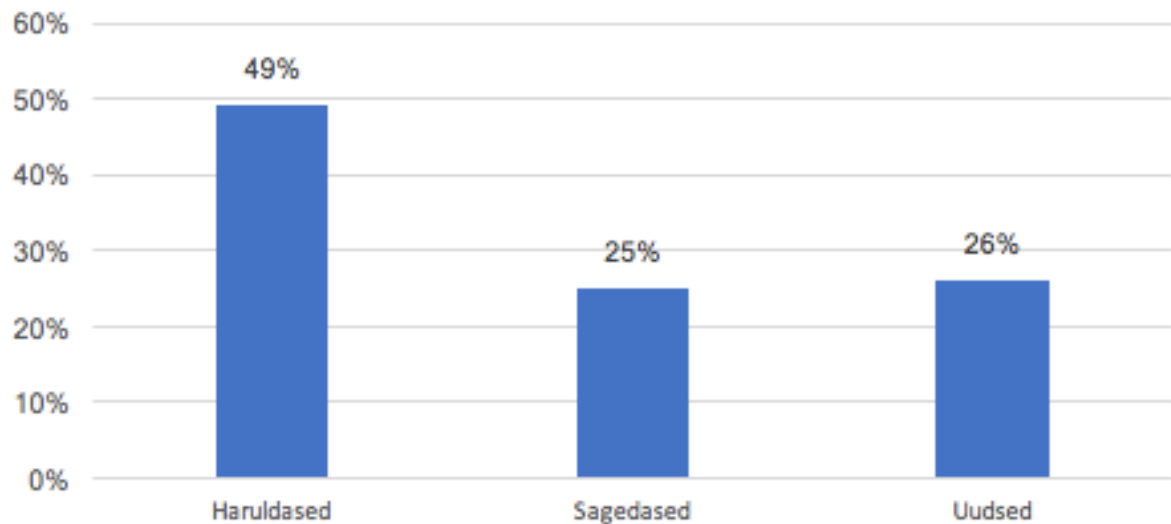


Joonis 4. SNVde jagunemine pre-miRNAde. X-teljel SNVde arv ning Y-teljel kõrgusaldusväärsusega pre-miRNAde arv.

Üle ühe SNV on 53-s pre-miRNAs, mis moodustab 35% kõigist pre-miRNAdest. Samas mitte üheski pre-miRNA ei olnud üle viie SNV, mis tõendab nende geneetiliste elementide olulist funktsiooni. Enim, ehk viis SNVd leidis kahes pre-miRNA. Need asusid seitsmendas kromosoomis miRNA miR-182-s ning 14 kromosoomis asuvas miR-431-s.

MiR-182-s on 109 nukleotiidses pre-miRNA viis SNVd. Kaks neist asuvad 3'pre-miRNA ning 3 juuksenõela struktuuris. Teine viie SNVga pre-miRNA asub 14 kromosoomis ning on miR-431. Seal asub üks SNV 5'pre-miRNA, üks juuksenõela struktuuris ning kolm 3'pre-miRNA.

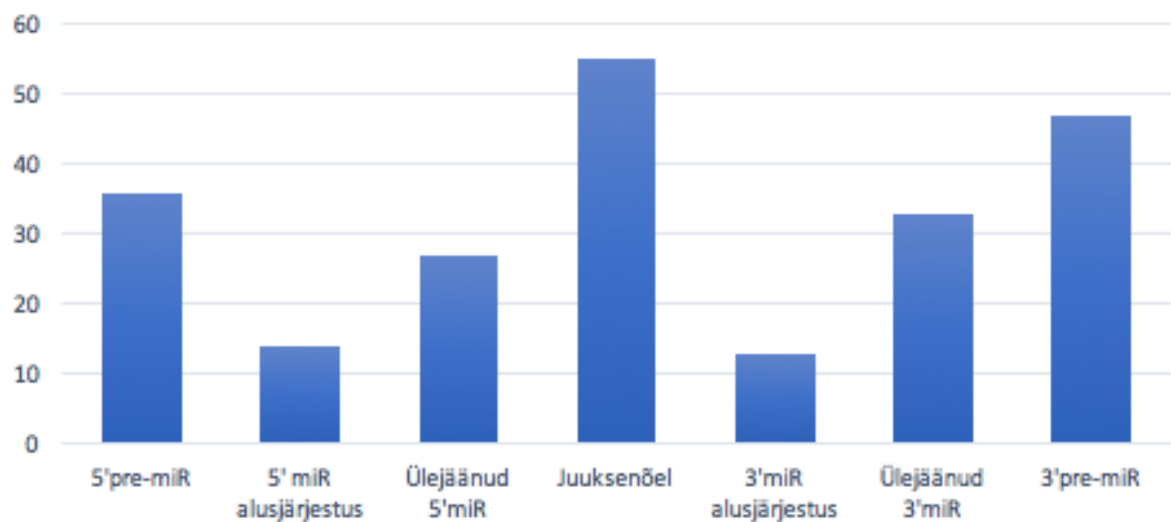
SNVde taoline jaotumine viitab miRNAde olulisusele, kuna siiski enamus neist (65%) ei oma ühtegi SNVd, mis tagab nendele miRNAdele ka võimaluse talitleda vastavalt enda algsele funktsioonile.



Joonis 5. SNVde jaotus esinemissageduse järgi. X-teljel SNVde jaotus vastavalt nende esinemissagedusele; Y-teljel on toodud protsentsuaalselt antud sagedusega SNVde esinemine EGV valimis.

Antud valimis on 1735 SNVd, millest 1288-t on varasemalt UCSC genoomi andmebaasis kirjeldatud. Sellest 1288st 853 on haruldased variandid ($AF < 1\%$) ning 435 sagedased variandid ($AF \geq 1\%$). 447 varianti ei ole varem kirjeldatud ning kategoriseeruvad seetõttu uudsete variantide alla.

Järgnevalt on vaadatud erineva sagedusega SNVde hulka miRNAdes vastavalt nende funktsionaalsele piirkonnale ja seda just kõrge usaldusväärsusega andmestikus võttes arvesse varasemalt välja toodud prekursor miRNA struktuuri jaotust. SNV leidis 144 pre-miRNAs.

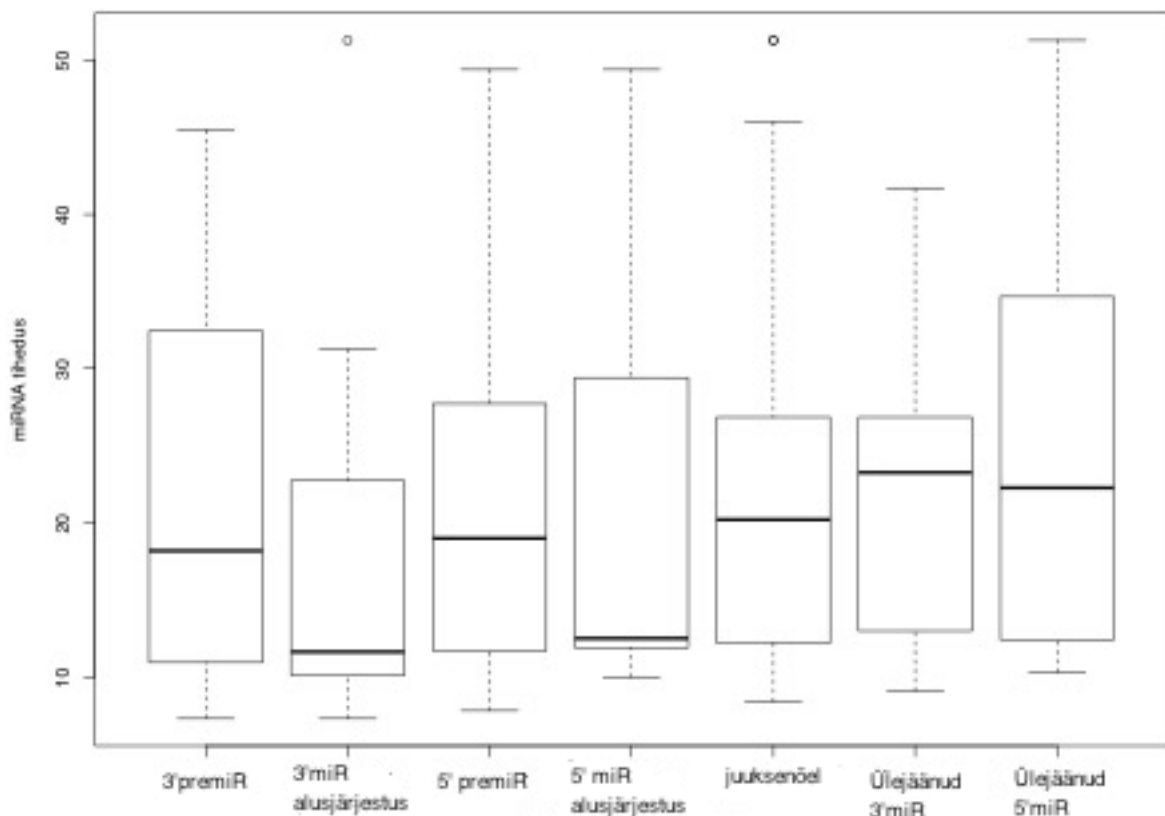


Joonis 6. Kõikide variantide jaotus pre-miRNA erinevates funktsionaalsetes osades. X-teljel on välja toodud pre-miRNA funktsionaalsed osad vastavalt varasemalt defineeritule; Y-teljel on näidatud vastavates pre-miRNA osades asuvate SNVde arv.

Vaadeldes SNVde jaotumist pre-miRNAs vastavalt miRNA funktsionaalsetele osadele näeme, et suurim hulk SNVsid leidub juuksenõela struktuuris, mis on ka oodatav, sest selle struktuuri osa miRNA funktsionaalsuse määramisel ei ole nii oluline kui näiteks alusjärjestuses, kus näeme ka varieeruvust kõige vähem.

27 SNVd mõjutavad miRNA alusjärjestusi, mis on olulised seondumisel miRNA sihtmärgiga ning seega ka miRNA funktsioneerimisel.

Erinevates miRNAdes ja nende osades on erineva hulga ning tihedusega mutatsioone. Selleks, et teha kindlaks nende mutatsioonide ulatus erinevates miRNA funktsionaalsetes osades vaadeldi SNVde tihedust.



Joonis 7. SNVde tihedus erinevates miRNA funktsionaalsetes osades. X-teljel on toodud erinevad miRNA funktsionaalsed osad ning y-teljel SNVde tihedus. Välja on toodud 3'premiR, 3'miR alusjärjestus, 5' premiR, 5'miR alusjärjestus, juuksenõela struktuur, ülejäänud 3' miRNA osa ning ülejäänud 5' miRNA osa. Karpdiagrammi keskmine joon toob esile mediaani. Ülemine piir esindab kolmandat kvartiili ning alumine piir esimest. Karpdiagrammi "vurrud" esindavad 95% usaldusintervalli maksimumi ja miinimumi. Ringid kujutavad erindeid, mis sellest intervallist välja jäävad.

Näeme jooniselt 7, et SNVde tiheduste mediaan on alusjärjestustes tunduvalt madalam kui ülejäänud osades. See on kooskõlas miRNA funktsioneerimise põhitõdedega, mis toob esile, et alusjärjestus on miRNA funktsioneerimise jaoks kriitilise olulisusega, kuna on tähtis seondumiseks mRNAle. Siiski näeme alusjärjestuses enim kõrvalekallet normaaljaotusest, kus mõne miRNA puhul saame täheldada ka mediaanist tunduvalt suuremat SNVde arvu. See tõendab asjaolu, et kuigi enamuses leidub miRNA alusjärjestuses vähesel määral SNVsid, on seal siiski mõne üksiku miRNA puhul SNVde tihedus suur.

2.3.2 Haigusseoselised SNVd miRNAdes

Variandid, mis mõjutavad miRNA toimimiseks eluvajalikke regioone, nagu nende alusjärjestus või juuksenõela struktuur, võivad kaasa tuua suure tõenäosusega tagajärgi miRNA funktsionaalsusele ning organismi fenotüübile.

Tabelites 2 ja 3 on toodud EGV valimis leitud SNVd, mis asuvad funktsionaalselt olulises alusjärjestuses. Kahte alleeli saab pidada sagedaseks, sest nende alleelisagedus ületab 1% piiri, rs2910164 (AF = 0.801916) ning rs2273626 (AF = 0.550802). Kokku asus funktsionaalselt olulises alusjärjestuses 24 SNVd, mis moodustab kogu kõrgeusaldusväärsusega pre-miRNAdes asuvatest SNVdest 10,7%.

Esimene sagedane alleel rs2910164 asub miR-146a-s, mida on varasemalt seostatud 143 erineva haigusega. Esile võib seal hulgas tõsta ühe miR-146-3p sihtmärgi PTTG1, mille valgu tase on positiivses korrelatsioonis põievähi migratsiooni, invasiooni, metastaasise ning kasvuga. MiR-146-3p üleekspressiooni puhul aga olid eelmainitud protsessid põievähi kontekstis inhibeeritud.

Teist pre-miRNA-t, miR-4707, mille alusjärjestuses asub sage alleel rs2273626 on seostatud kroonilise obstruktiivse kopsuhaigusega (Wang, et al., 2016).

Seega SNV antud miRNAde alusjärjestuses võib mõjutada antud miRNA seondumist oma sihtmärk geenile, mille tagajärjel võib miRNA seondumine olla häiritud, mis omakorda võib viia fenotüübiliste tulemusteni.

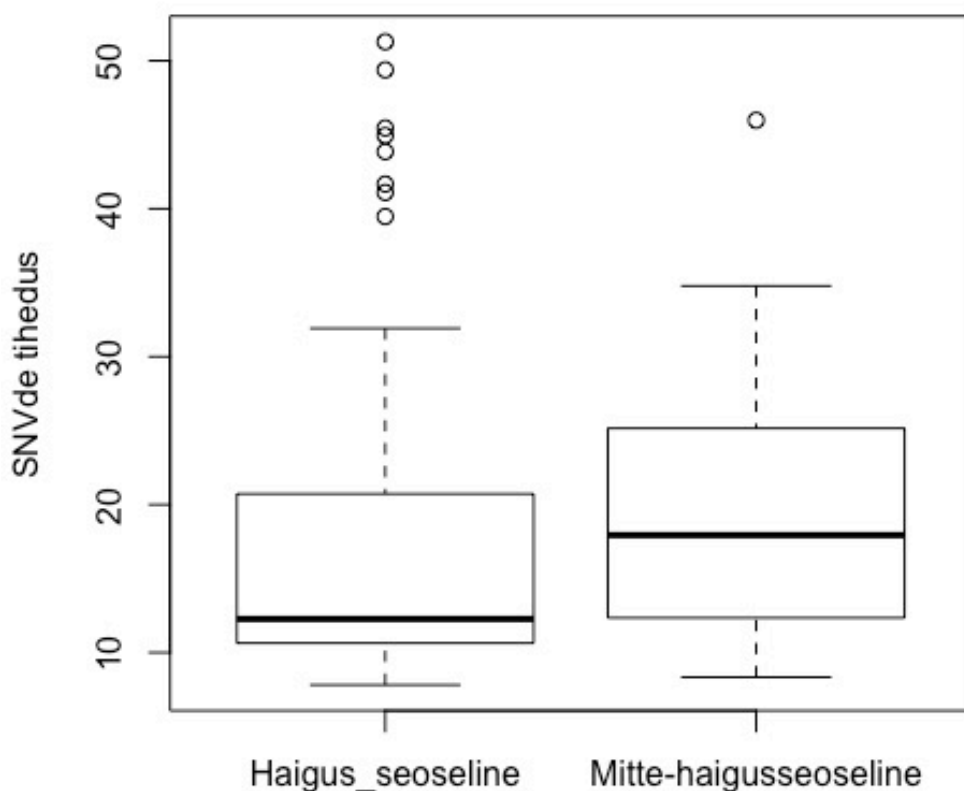
Tabel 2. MiRNAd 3' alusjärjestuses asetsevad SNVd. pre-miRNA- premiRNA, kus antud SNV leidub; SNV– referents ning mutatsiooniga alleel, variandi number dbSNP andmebaasis (olemasolul); asukoht – SNV kromosomaalne asukoht; AF- alleelisagedus, millega antud variant EGV valimis esineb; haigus(ed) – haigused, millega antud miRNAd kirjanduse alusel seostatud on.

| pre-miRNA | SNV | Asukoht | AF | Haigus(ed) |
|-------------------|-------------------|-----------------|-------------|---|
| miR-146a | C/G (rs2910164) | Chr5:159912418 | 0.801916 | Pöievähk (Xiang <i>et al.</i> , 2017) |
| miR-1296 | C/T | Chr10:65132745 | 0.000222816 | Maovähk (Shan <i>et al.</i> , 2017) |
| miR-192 | T/C | Chr 11:64658647 | 0.000891266 | Kopsuvähk (Mancuso <i>et al.</i> , 2016) |
| miR-493 | G/T | Chr14:101335457 | 0.000222816 | Prolaktinoom (Chen <i>et al.</i> , 2012) |
| miR-323a | G/A | Chr14:101492118 | 0.000222816 | Epilepsia (Che <i>et al.</i> , 2017) |
| miR-181d | G/A (rs369928561) | Chr19: 13985772 | 0.000222816 | Glioblastoom (Zhang <i>et al.</i> , 2012) |
| miR-1249 | G/A | Chr22: 45596859 | 0.000222816 | Melanoom (Leidinger <i>et al.</i> , 2010) |
| miR-4536-1 | A/G | ChrX: 55477948 | 0.000222816 | - |
| miR-615 | G/A | Chr12: 54427797 | 0.000222816 | Maksavähk (Wu <i>et al.</i> , 2016) |
| miR-96 | A/G | Chr7: 129414553 | 0.000222816 | Pöievähk (Eissa <i>et al.</i> , 2015) |
| miR-1307 | G/A | Chr10:105154073 | 0.000222816 | Eesnäärmevähk (Qiu ja Dou, 2017) |
| miR-504 | C/T | ChrX: 137749901 | 0.000445633 | Glioom (Guan <i>et al.</i> , 2015) |
| miR-506 | G/A | ChrX: 146312284 | 0.000222816 | Kõhunäärmevähk (Cheng <i>et al.</i> , 2016) |

Tabel 3. MiRNAde 5' alusjärjestuses asetsevad SNVd. pre-miRNA- premiRNA, kus antud SNV leidub; SNV– referents ning mutatsiooniga alleel, variandi number dbSNP andmebaasis (olemasolul); Asukoht – SNV kromosomaalne asukoht; AF- alleelisagedus, millega antud variant EGV valimis esineb; haigus(ed) – haigused, millega antud miRNAd kirjanduse alusel seostatud on.

| pre-miRNA | SNV | Asukoht | AF | Haigus(ed) |
|-------------------|-------------------|-------------------|-------------|---|
| miR-488 | C/T (rs199722070) | Chr1: 176998564 | 0.000891266 | Maovähk (Wu <i>et al.</i> , 2016) |
| miR-128-2 | G/A | Chr3: 35785986 | 0.000445633 | Hepatotsellulaarne kartsinoom (Zhuang <i>et al.</i> , 2015) |
| miR-25 | C/T | Chr7: 99691249 | 0.000222816 | Kõhunäärmevähk ((Deng, et al., 2016) |
| let-7a-1 | T/A (rs201607747) | Chr9: 96938244 | 0.000445633 | Ajukasvaja (Li, Zhang, Chen, & Ma, 2016) |
| miR-450a-2 | A/G (rs141784874) | ChrX: 133674612 | 0.00557041 | - |
| miR-4707 | C/A (rs2273626) | Chr14: 23426182 | 0.550802 | - |
| miR-433 | C/T | Chr14: 101348236 | 0.000222816 | Müeloproliferatiivne kasvaja (Lin, et al., 2013) |
| miR-302c | T/C | Chr 4 : 113569575 | 0.000222816 | Ajukasvaja ((Wang, et al., 2015) |
| miR-627 | A/G | Chr15:42491772 | 0.000222816 | Kõhunäärmevähk ((Shin, Ng, Chan, Kwong, & Chu, 2015) |
| miR-328 | C/G (rs188892061) | Chr16 : 67236292 | 0.00356506 | Melanoom (Li, Wang, Gong, Fang, & Guo, 2015) |
| miR-654 | A/T | Chr14: 101506570 | 0.0013369 | Eesnäärmevähk (Formosa, et al., 2014) |

HMDD (*Human microRNA Disease Database*) on käsitsi kogutud miRNAde ja haiguste seoselisuse andmebaas. Võrreldes seda kõrge usaldusväärsusega pre-miRNA andmebaasiga, täheldasime et 295 pre-miRNAst 245 (83%) on varem seostatud mõne haigusega ning 50 (17%) mitte. MiRNAat, mis on assotseerunud vähemalt ühe haigusega nimetatakse haigusseoseliseks miRNAks ning kui ühtegi seost ühegi haigusega varem täheldatud ei ole, nimetatakse miRNAat mitte-haigusseoseliseks.



Joonis 8. SNVde tihedus vastavalt haigus- ning mitte-haigusseoselistes miRNAdeks. X teljel on toodud haigus ning mitte-haigusseoselised miRNAde ning Y-teljel SNVde tihedus. Karpdiagrammi keskmine joon toob esile mediaani. Ülemine piir esindab kolmandat kvartiili ning alumine piir esimest. Karpdiagrammi "vurrud" esindavad 95% usaldusintervalli maksimumi ja miinimumi. Ringid kujutavad erindeid, mis sellest intervallist välja jäävad.

Joonis 5 näitab, et SNVde tihedus haigus-seoselistes miRNAdes on madalam kui mitte-haigusseoselistes miRNAdes, kuid see tulemus ei ole statistiliselt oluline (t-test; p-väärtus – 0.2294). Kõrge p-väärtus võib tuleneda ka antud valimi väiksusest. Siiski on see ka varasemale informatsioonile toetudes oodatav tulemus.

Huvitavaks asjaoluks antud tulemuste juures võib pidada aga seda, et jaotus mitte-haigusseoseliste puhul on lähemal normaaljaotusele kui haigusseoseliste puhul. Ehk isegi kui keskmiselt leiame haigus-seoselistest miRNAdest vähem SNVsid, kui mitte-haigusseoselistest, siis erindeid on haigus-seoseliste puhul tunduvalt rohkem. Mis annab omakorda aga kinnitust asjaolule, et osade haigus-seoseliste miRNAde puhul leiame SNVsid vägagi suure tihedusega.

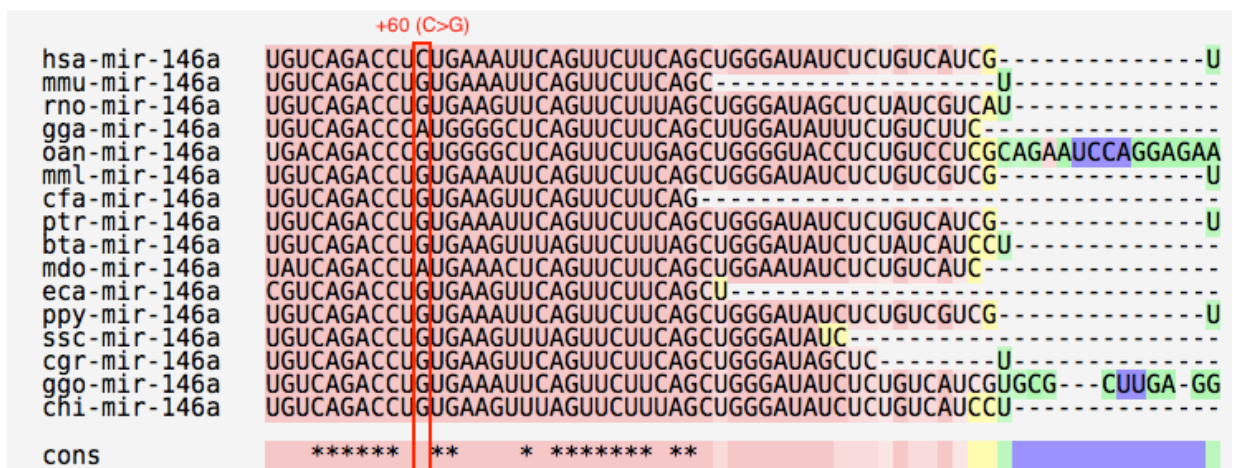
2.3.3 miRNA alusjärjestuses paiknevad SNVd evolutsioonilise konserveerumise taustal

Kuna paljud SNVd esinevad antud valimis kõrge alleelisagedusega on huvitav kindlaks teha ka antud populatsioonis leitud positsioonide konserveerumine läbi liikide ning seda just selgroogsete hulgas. Järgnevalt ongi kahe miRNA (mir-146a ja mir-654) alusjärjestuses asuva SNV puhul vaadeldud evolutsioonilist konserveeritust.

Mir-146a-s paikneb SNV (rs2910164) positsioonil +60 (C>G). Allpool toodud joonduselt näeme, et tegu on inimese spetsiifilise muutusega.

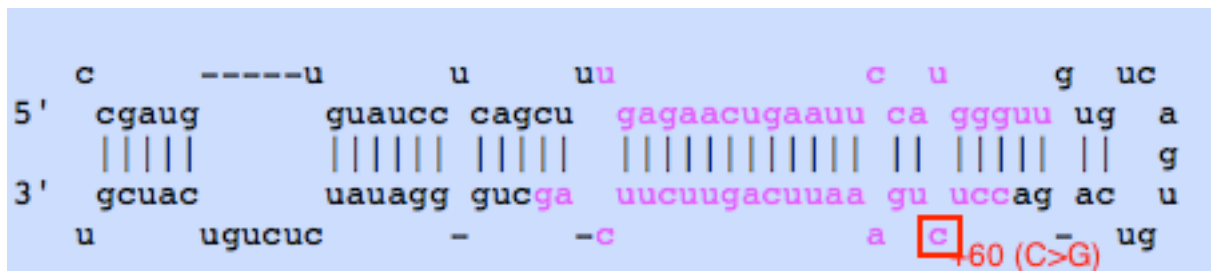
Antud SNV asub miRNA 3' alusjärjestuses ning selle miRNA ekspressiooni on enim täheldatud rinnakoes, vereplasmas ning maksa kudedes (*Human miRNA expression database*). Enim on antud miRNA tuvastatud just rinnavähiga seotud kudedes. Patsientidel, kel esines rinna- või munasarjavähk, diagnoositi haigus varem just neil, kellel esines antud lookuses C alleel võrreldes nendega, kel esines G alleel (Shen *et al.*, 2008).

Huvitavaks teeb antud SNV ka asjaolu, et tegemist on inimese spetsiifilise variatsiooniga.



Joonis 9. Mitmene joondus pre-mir-146a järjestusele segroogsete hulgas, anoteeritud hsa (*Homo Sapiens*) järjestuse järgi.

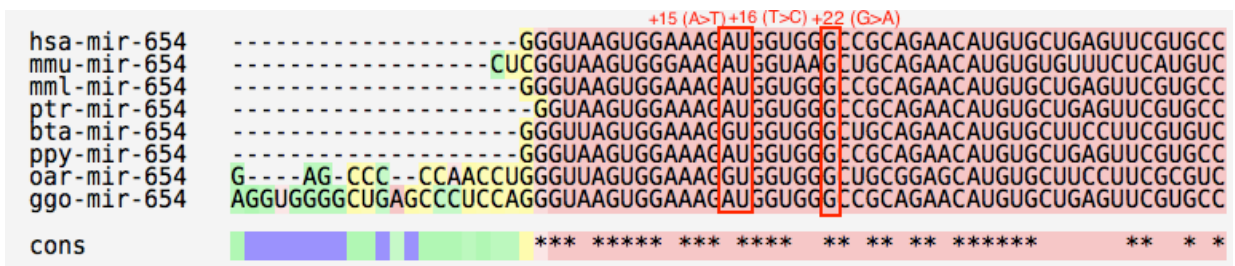
Hsa, *Homo sapiens*; mmu, *Mus musculus*; rno, *Rattus norvegicus*; gga, *Gallus gallus*; oan, *Ornithorhynchus anatinus*; mml, *Macaca mulatta*; cfa, *Canis familiaris*; ptr, *Pan troglodytes*; bta, *Bos taurus*; mdo, *Monodelphis domestica*; eca, *Equus caballus*; ppy, *Pongo pygmaeus*; ssc, *Sus scrofa*; cgr, *Cricetulus griseus*; ggo, *Gorilla gorilla*; chi, *Capra hircus*.



Joonis 10. Pre-mir-146a juuksenõela struktuur mirBASE'ist. Esile on toodud 60 positsioonil olev SNV C>G. Roosaga on kujutatud miRNA asukoht, must on pre-miRNA ning juuksenõela struktuur.

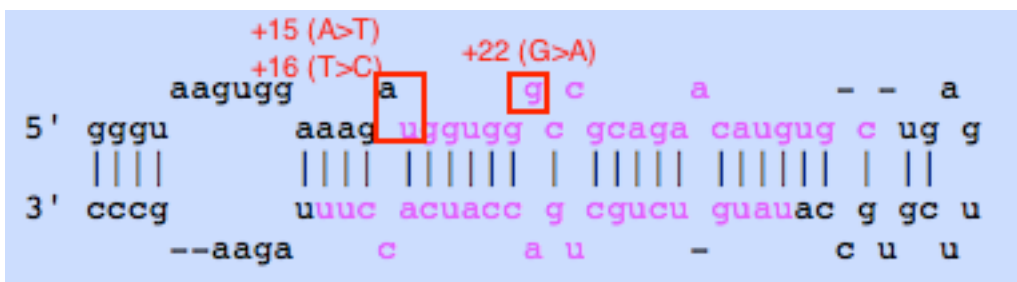
Teine miRNA, mille alusjärjestuses leitud SNVsid on mir-654, mille 5' alusjärjestuses leiti EGV valimis 3 SNVd positsioonidel +15 (A< T), +16 (T<C) ning +22 (G<A).

Antud miRNA ekspressiooni on enim täheldatud rinna kudedes, vereplasmas ning pankrease kudedes (*Human miRNA expression database*). Antud miRNA puhul näeme, referentsalleelide puhul konserveerumist läbi liikide.



Joonis 11. Mitmene joondus pre-mir-654 järjestusele selgroogsete hulgas miRBASE’is leiduvatele järjestustele. Annoteeritud hsa (*Homo sapiens*) järjestuste järgi.

Hsa, *Homo sapiens*; mmu, *Mus musculus*; mml, *Macaca mulatta*; ptr, *Pan troglodytes*; bta, *Bos taurus*; ppy, *Pongo pygmaeus*; ssc, *Sus scrofa*; oar, *Ovis aries*; ggo, *Gorilla gorilla*.



Joonis 12. Pre-mir-654 juuksenõela struktuur mirBASE’ist. Punase kastiga on esile toodud 15,16 ning 22 positsioonil asuvad SNVd. Roosaga on kujutatud miRNA asukoht, must on pre-miRNA ning juuksenõela struktuur.

Nagu näeme on osad variandid konserveerunud läbi kõikide selgroogsete ning osade puhul sellist trendi täheldada ei saa.

Antud tulemustest nägime, et EGV valimis leidis pre-miRNAdes SNVsid enim juuksenõela struktuuris ning vähim miRNA alusjärjestustes. Samuti kindlasti on märkimisväärne alusjärjestuses SNVd omavate miRNAdes haigusseoselises, seetõttu nägime, et vaid 2 miRNAt, mille alusjärjestuses on SNV, ei ole hetkel veel seostatud ühegi haigusega. See annab kinnitust, et miRNAdel on ka oluline osa haigus tekkes ning eriti võib see mõjutatud olla variatsioonidest nende järjestustes.

KOKKUVÕTE

Avastuste arv, mis seovad miRNAsid erinevate haigustega kasvab igapäevaselt ning seega on ülimalt oluline mõista miRNA erinevate funktsionaalsete osade konserveerumise taset ning variatsioone nende geenides, seda just eesmärgiga teha kindlaks patoloogilised variandid.

Antud töös näeme pre-miRNA struktuuris (joonis 5) SNVde jaotumise kooskõla tema funktsionaalsete osade olulisusega. Mistõttu enim konserveerunud on miRNA alusjärjestus ning enim variatsioone leiame juuksenõela struktuuris. Samuti on SNVde tihedus funktsionaalsuse jaoks olulisemates regioonides madalam kui teistes miRNA regioonides.

Samamoodi sai ka kinnitust asjaolu, et haigus-seoslistes miRNAdes on ka EGV variantide SNVde tihedus madalam kui mitte-haigusseoslistes. Siiski leidub haigus-seoselistes enim erindeid, mis viitab mõnede miRNAde puhul väga paljudele SNVdele üpriski lühikestes järjestustes.

Vaadeldes evolutsioonilist konserveerumist kahe miRNA näitel, täheldasime ühel juhul SNVsid läbi selgroogsete (kaasa arvatud inimene) konserveerunud positsioonis, kuid teise puhul nägime variatsiooni positsioonis, mis on konserveerunud kõikides vaadeldud selgroogsetes, kuid erinev inimese genoomis.

Siiski seostatakse variatsioonidega miRNAsid enim erinevate haigustega, eelkõige vähiga, mis viitab kindlasti asjaolule, et variatsioonid on olulised miRNAde haigusseoselisuse juures.

Edaspidi on kindlasti oluline võrrelda varieeruvust ka teiste populatsioonidega ning teha kindlaks, kas on mõningaid variatsioone, mis on omased vaid Eesti populatsioonile.

Variation in microRNA genes

Mirjam Tamme

SUMMARY

MicroRNAs (miRNAs) are small non-coding RNAs, which are important factors in gene regulation. They are relevant in all main cell processes, including cell division, differentiation and apoptosis. Considering the importance of single nucleotide variations (SNVs) they can have an impact on the function of those small molecules and to the functioning of human organism.

In this thesis we observed the variations in miRNA genes based on the Estonian Genome Center dataset. The size of the sample was 2244 individuals, from where we found 1735 SNVs, from which 1288 had already earlier been described in the dbSNP database. The miRNA data was received from miRBASE database, where we had 1877 human pre-miRNAs, from which 296 were in the high-confidence dataset and 2810 miRNAs, from which 617 belonged to the high-confidence dataset. We observed that most of the miRNAs had no SNVs in their sequence, over 1 SNV was present in 53 pre-miRNAs, which makes 35% of all the pre-miRNAs. Still no miRNA possessed more than 5 SNVs, which shows the importance of these genetic elements.

We concluded that regions, which are more important in terms of binding to mRNA (seed regions) are more conserved than those functional parts of a miRNA, which are not so important (loop region).

Since miRNAs and especially variation in different miRNAs is often related to different human diseases, we also observed miRNA-s which possessed a SNV in their seed region and their relation to different diseases, which made clear that nearly all miRNAs which possessed a SNV in their seed region are also previously reported to be related to at least one disease. Finally we observed two miRNAs and SNVs in their seed region considering also their evolutionary conservation, from which we were able to conclude that one of the SNVs was only human specific, when the other was present in all observed vertebrates.

As further research it would be beneficial to compare these variations with other populations and investigate whether there are some variations that only occur in the Estonian population.

TÄNUSÕNAD

Täna antud töö valmimise juures enda juhendajat, Tarmo Annilot, kannatlikkuse ning väga asjalike kommentaaride ning paranduste eest!

3. KASUTATUD KIRJANDUS

- Allen, E., Xie, Z., Gustafson, A., Sung, G., Spatafora, J., & Carrington, J. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet*, 1282-1290.
- Altuvia, Y., Landgraf, P., Lihtwick, G., Elefant, N., Pfeffer, S., Aravin, A., . . . Margalit, H. (2005). Clustering and conservation patterns of human microRNAs. *Nucleic Acids*, 2697-2706.
- Bartel, D. P. (2004, Jaanuar). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, 116, 281-297.
- Baskerville, S., & Bartel, D. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, 241-247.
- Bhartiya, D., Saurabh, L., Mukhopadhyay, A., & Scaria, V. (2011). miRvar: A comprehensive database for genomic variations in microRNAs. *Human Mutation*, 2226-2245.
- Cammaerts, S., Strazisar, M., Dierckx, J., Del Favero, J., & De Rijk, P. (2015). miRVaS: a tool to predict the impact on genetic variants on miRNAs. *Nucl. Acids*, 1093.
- Carbonell, J., Alloza, E., Arce, P., Borrego, S., Santoyo, J., Ruiz-Ferrer, M., . . . Dopazo, J. (2012). A map on human microRNA variation uncovers unexpectedly high levels of variability. *GenomeMedicine*, 4:62.
- Carthew, R., & Sontheimer, E. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 642-55.
- Che, N., Zu, G., Zhou, T., Wang, X., Sun, Y., Tan, Z., . . . Yin, J. (2017). Aberrant Expression of miR-323a-5p in Patients with Refractory Epilepsy Caused by Focal Cortical Dysplasia. *Genet Test Mol Biomarkers*, 3-9.
- Chen, C., Li, L., Lodish, H., & Bartel, D. (2004, Septermber). MicroRNAs modulate hematopoietic lineage differentiation. *Science*, 83-86.
- Chen, Y., Li, Q., Wang, C., Su, Z., Li, W., Chen, X., & Wu, Z. (2012). Differential expression analysis of prolactinoma-related microRNAs. *Zhonghua Yi Xue Za Zhi*, 320-323.
- Cheng, R., Wang, J., Zhang, J., Sun, L., Zhao, Y., Qiu, Z., . . . Sun, Y. (2016). MicroRNA-506 is up-regulated in the development of pancreatic ductal adenocarcinoma and is associated with attenuated disease progression. *Chin J Cancer*.
- Chi, S., Zang, J., Mele, A., & Darnell, R. (2009). Ago HITS-CLIP decodes miRNA-mRNA interaction maps. *Nature*, 479-486.
- Cuperus, J., Fahlgren, N., & Carrington, J. (2011). Evolution and functional diversification of miRNA genes. *Plant Cell*, 431-442.

- Deng, T., Yuan, Y., Zhang, C., Yao, W., Wang, C., Liu, R., & Ba, Y. (2016). Identification of Circulating MiR-25 as a Potential Biomarker for Pancreatic Cancer Diagnosis. *Cell Physiol Biochem*, 1716-1722.
- DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., Hartl, C., . . . Daly, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 491-498.
- Dweep, H., Georgiou, G. D., Gretz, N., Deltas, C., Voskarides, K., & Felekis, K. (2013). CNVs-microRNAs Interactions Demonstrate Unique Characteristics in the Human Genome. An Interspecies In silico Analysis. *PLOS one*.
- Edeleva, E., & Shcherbata, H. (2013). Stress-induced ECM alteration modulates cellular microRNAs that feedback to readjust the extracellular environment and cell behavior. *Front Genet*, 305.
- Eissa, S., Habib, H., Ali, E., & Kotb, Y. (2015). Evaluation of urinary miRNA-96 as a potential biomarker for bladder cancer diagnosis. *Medical Oncology*.
- Felekis, K., Voskarides, K., Dweep, H., Sticht, C., & Gretz, N. (2011). Increased number of microRNA target sites in genes encoded in CNV regions. Evidence for an evolutionary genomic interaction. *Mol Biol Evol*, 2421-2424.
- Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008, February). Mechanism of post-transcriptional regulation by microRNAs: are the answers in sight ? *Nature Reviews Genetics*, 102-14.
- Formosa, A., Markert, E., Lena, A., Italiano, D., Finazzi-Agro, E., Levine, A., . . . Candi, E. (2014). MicroRNAs, miR-154, miR-299-5p, miR-376a, miR-376c, miR-377, miR-381, miR-487b, miR-485-3p, miR-495 and miR-654-3p, mapped to the 14q32.31 locus, regulate proliferation, apoptosis, migration and invasion in metastatic prostate cancer cells. 5173-51782.
- Gamazon, E., Ziliak, D., Im, H., LaCroix, B., Park, D., & Cox, N. (2012). Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. *Human Genetics*, 1046-1063.
- Gong, J., Liu, C., Wu, Y., Ma, Z., Chen, H., & Guo, A. (2015). An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Oxford Academic*, 1093.
- Gong, J., Tong, Y., Zhang, H.-M., Wang, K., Hu, T., Shan, G., . . . Guo, A.-Y. (2011). Genome-Wide Identification of SNPs in MicroRNA Genes and the SNP Effects on MicroRNA Target Binding and Biogenesis. *Human Genome Variation Society*, 254-263.

- Griffiths-Jones, S., Saini, H., van Dongen, S., & Enright, A. (2008). mirBASE:tools for microRNA genomics. *Nucl Acids Res*, 154-158.
- Guan, Y., Chen, L., Bao, Y., Pang, C., Cui, R., Li, G., . . . Wang, Y. (2015). Downregulation of microRNA-504 is associated with poor prognosis in high-grade glioma. *Int J Clin Exp Pathol*, 727-734.
- Guo, M. H., Nandakumar, S. K., Ulirsch, J. C., Zekavat, S. M., Buenrosto, J. D., Natarajan, P., . . . Sankaran, V. G. (2017). Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. *PNAS*, 327-336.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Berninger, P., & Rothballer, A. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 129-141.
- Hariharan, M., Scaria, V., & Brachmachari, S. (2009). dbSMR: a novel resource of genome-wide SNPs affecting microRNA mediated regulation. *BMC Bioinformatics*, 10:108.
- Heimberg, A., Sempere, L., Moy, V., Donoghue, P., & Peterson, K. (2008). MicroRNAs and the advent of vertebrate morphological complexity. 2946-2950.
- Helwak, A., Kudla, G., Dudankova, T., & Tollervy, D. (2013). Mapping the human mirNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 654-665.
- Hogg, D., & Harries, L. (2014). Human genetic variants and its effect on miRNA biogenesis, activity and function. *Bichem*, 1184-1189.
- Horikawa, Y., Wood, C., Yang, H., Zhao, H., Ye, Y., Gu, J., . . . Wu, X. (2008). Single Nucleotide polymorphisms of microRNA processing machinery genes modify the risk of renal cell carcinoma. *Clin.Cancer*, 7956-7962.
- Hrdlickova, B., Coutinho de Almeida, R., Borek, Z., & Withoff, S. (2014). Genetic variation in the non-coding genome:Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica* , 1910-1922.
- Iafrate, A., Feuk, I., Rivera, M., Listewnik, M., & Donahoe, P. (2004). Detection of large-scale variation in the human genome. *Nat Genet*, 949-951.
- Jazdzewski, K., Murray, E., Franssila, K., Jarzab, B., Schoeneberg, D., & de la Chapelle, A. (2008). Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc natl Acad Sci*, 7269-7274.
- Johnson, S., Lin, S., & Slack, F. (2003). The time of appearance of the C.elegans let-7 microRNA is transcriptionally controlled utilizing atemporal regulatory element in its promoter. *Dev. Biol.*, 364-379.

- Jung, Y., Jeon, Y., Rah, H., Kim, J., Shin, J., Choi, D., . . . Kim, N. (2014). Genetic variants in microRNA machinery genes are associated with idiopathic recurrent pregnancy loss risk. *PLoS one*, e95803.
- Karolchik, D., Hinrichs, A. S., Furey, T. S., & Kent, J. W. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, D493-D496.
- Kent, W., Sugnet, C., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human Genome Browser at UCSC. *Genome Research*, 996-1006.
- Kim, H. K., Prokunina-Olsson, L., & Chanock, S. (2012). Common genetic Variants in miR-1206(8q24.2) and miR-612(11q13.3) Affect Biogenesis of Mature miRNA Forms. *PLoS One*, 7-10.
- Kloosterman, W., & Plasterk, R. (2006). The diverse functions of microRNAs in animal development and disease. *Dev Cell*, 441-450.
- Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., & Tuschl, T. (2003). New microRNAs from mouse and human. *RNA*, 175-179.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., & Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Curr Biol*, 735-739.
- Lee Y, J. K. (2002, Sep). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO Journal*, 4663-70.
- Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 843-54.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., . . . Kim, V. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 415-419.
- Leidinger, P., Keller, A., Borries, A., Reichrath, J., Rass, K., Jager, S. U., . . . Meese, E. (2010). High-throughput miRNA profiling of human melanoma blood samples. *BMC Cancer*.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 1754-1760.
- Li, J., Wang, J., Gong, Q., Fang, R., & Guo, Y. (2015). MicroRNA-328 inhibits proliferation of human melanoma cells by targeting TGF β 2. *Asian Pac J Cancer Prev*, 1575-1579.
- Li, Y., Zhang, X., Chen, D., & Ma, C. (2016). Let-7a suppresses glioma cell proliferation and invasion through TGF- β /Smad3 signaling pathway by targeting HMGA2. *Tumour Biol*, 8107-8119.
- Lin, X., Rice, K., Buzzai, M., Hexner, E., Costa, F., Kilpivaara, O., . . . Licht, J. (2013). miR-433 is aberrantly expressed in myeloproliferative neoplasms and suppresses hematopoietic cell growth and differentiation. *Leukemia*, 344-352.

- Liu, C., Zhang, F., Tingting, L., Lu, M., Wang, L., Yue, W., & Zhang, D. (2012). MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics*, 13:661.
- Liu, S., An, J., Lin, J., Liu, Y., Bao, L., Zhang, W., & Zhao, J. (2014). Single Nucleotide polymorphisms of microRNA processing machinery genes and outcome of hepatocellular carcinoma. *PLoS one*, e92791.
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., & Cui, Q. (2008). An analysis of human microRNA and disease associations. *PLoS ONE*, e3420.
- Lukiw, W. J., Zhao, Y., & Cui, J. G. (2008). An NF- κ B-sensitive Micro RNA-146a-mediated Inflammatory Circuit in Alzheimer Disease and in Stressed Human Brain Cells. *Journal of biological Chemistry*, 31315–31322.
- Mancuso, G., Bovio, E., Rena, O., Rrapaj, E., Mercalli, F., Veggiani, C., . . . Boldorini, R. (2016). Prognostic impact of a 3-MicroRNA signature in cytological samples of small cell lung cancer. *Cancer Cytopathol*, 621-629.
- Mattick, J. S., & Makunin, I. V. (2005). Small regulatory RNAs in mammals. *Human Molecular Genetics*, 121-132.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., . . . DePristo, M. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 1297-1303.
- Megraw, M., Sethpathy, P., Corda, B., & Hatzigeorgiou, A. (2007). miRGen: a database for the study of animal microRNA genomic organization and function. *Nucleic Acids*, 149-155.
- Mencia, A., Modamio-Hoybjor, S., Redshaw, N., Morin, M., Mayo-Merino, F., Olavarriata, L., . . . Dalmay, T. (2009). Mutations in the seed region of human mir-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet*, 609-613.
- Mu, Y., & Su, X. (2012). Polymorphism in pre-miR-30c contributes to gastric cancer risk in a Chinese population. *Med Oncol*, 1723-1732.
- Nozawa, M., Miura, S., & Nei, M. (2010). Origins and Evolution of MicroRNA Genes in Drosophila Species. *Genome Biol Evol*, 180-189.
- Olena, A., & Patton, J. (2010). Genomic organization of microRNAs. *Cell Pyhsiology*, 540-545.
- Pashkovskiy, P., & Ryazansky, S. (2013). Biogenesis, evolutsion, and functions of plant microRNAs. *Biochemistry*, 627-637.

- Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., . . . Ruvkun, G. (2000, November). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 86-89.
- Permuth-Wey, J., Thompson, R., Burton, L. N., Olson, J., Browning, J., Madden, M., . . . Egan, K. (2011). A functional polymorphism in the pre-miR-146a gene is associated with risk and prognosis in adult glioma. *J Neurooncol*, 639-646.
- Peterson, K., Dietrich, M., & McPeck, M. (2009). MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays*, 736-747.
- Qiu, X., & Dou, Y. (2017). miR-1307 promotes the proliferation of prostate cancer by targeting FOXO3A. *Biomed Pharmacother*, 430-435.
- Rah, H., Jeon, Y., Lee, B., Kim, J., Shim, S., Lee, W., . . . Kim, N. (2013). Association of polymorphisms in microRNA machinery genes (DROSHA, DICER1, RAN and XPO5) with risk of idiopathic primary ovarian insufficiency in Korean women. *Menopaus*, 1067-1073.
- Rajewsky, N. (2006). microRNA target predictions in animals. *Natural Genetics*, 38(Suppl)S:8-13.
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., & Bartel, D. P. (2000). MicroRNAs in plants. *Genes Dev*, 1616-1626.
- Ryan, B., Robles, A., & Harris, C. (2010). Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer*, 389-402.
- Shalgi, R., Lieber, D., Oren, M., & Pilpel, Y. (2007). Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol*, e131-10.
- Shan, X., Wen, W., Zhu, D., Yan, T., Cheng, W., Huang, Z., . . . Zhu, J. (2017). miR 1296-5p Inhibits the Migration and Invasion of Gastric Cancer Cells by Repressing ERBB2 Expression.
- Shastry, B. (2009). SNPs: impact on gene function and phenotype. *Mol Biol*, 3-22.
- Shen, J., Ambrosone, C., & Zhao, H. (2009). Novel genetic variants in microRNA genes and familial breast cancer. *Int.J.Cancer*, 1178-1182.
- Shen, J., Ambrosone, C., DiCioccio, R., Odunsi, K., Lele, S., & Zhao, H. (2008). A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis*, 1963-6.
- Shin, V., Ng, E., Chan, V., Kwong, A., & Chu, K. (2015). A three-miRNA signature as promising non-invasive diagnostic marker for gastric cancer. *Mol Cancer*.

- Sun, G., Yan, J., Noltner, K., Feng, J., Li, H., Sarkis, D., . . . Rossi, J. (2009). SNPs in human miRNA genes affect biogenesis and function. *RNA*, 1640-1651.
- Szymanski, M., Barciszewska, M. Z., Erdmann, V. A., & Berciszewski, J. (2005). A new frontier for molecular medicine: Noncoding RNAs. *Biochimica et Biophysica Acta* , 65-75.
- The International HapMap Consortium. (2003). International HapMap project. *Nature*, 789-796.
- Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J.-M., . . . Noterdame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*, W13-W17.
- Tumbull, C., Ahmed, S., Morrison, J., Pernet , D., Renwick, A., Maranian, M., . . . Easton, D. (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genetics*, 504-507.
- Tuna, M., Machado, A. S., & Calin, G. A. (2016). Genetic and Epigenetic Alterations of MicroRNAs and Implications for Human Cancers and Other Diseases. *Genes, Chromosomes & Cancer*, 193-214.
- Van der Auwera, G., Carneiro, M., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., . . . DelPristo, M. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 1-33.
- Wang, M., Huang, Y., Liang, Z., Liu, D., Lu, Y., Dai, Y., . . . Wang, C. (2016). Plasma miRNAs might be promising biomarkers of chronic obstructive pulmonary disease. *The Clinical Respiratory Journal*, 104-111.
- Wang, Y., Wei, Y., Tong, H., Chen, L., Fan, Y., Ji, Y., . . . Wang, G. (2015). MiR-302c-3p suppresses invasion and proliferation of glioma cells via down-regulating metadherin (MTDH) expression. *Cancer Biol Ther.*, 1308-1315.
- Warthmann, N., Das, S., Lanz, C., & Weigel, D. (2008). Comparative analysis of the MIR319a microRNA locus in Arabidopsis and related Brassicaceae. *Mol Biol Evol*, 892-902.
- Wheeler, B., Heimberg, A., Moy, V., Sperling, E., Holstein, T., Heber, S., & Peterson, K. (2009). The deep evolution on metazoan microRNAs. *Evol Dev*, 50-68.
- Wu, X., Deng, L., Tang, D., Ying, G., Yao, X., Liu, F., & Liang, G. (2016). miR-615-5p prevents proliferation and migration through negatively regulating serine hydromethyltransferase 2 (SHMT2) in hepatocellular carcinoma. *Tumor Biology*, 6813-6821.

- Wu, X., Tang, H., Liu, G., Wang, H., Shu, J., & Sun, F. (2016). miR-448 suppressed gastric cancer proliferation and invasion by regulating ADAM10. *Tumour Biol*, 10545-51.
- Xiang, W., Wu, X., Huang, C., Wang, M., Zhao, X., Luo, G., . . . Zeng, F. (2017). PTTG1 regulated by miR-146a-3p promotes bladder cancer migration, invasion, metastasis and growth. *Oncotarget*, 664-678.
- Yi, R., Qin, Y., Macara, I., & Cullen, B. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*, 3011-3016.
- Zeng, Y., & Cullen, B. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc.Natl.Acad.Sci*, 9779-9784.
- Zhang, W., Zhang, J., Hoadley, K., Kushwaha, D., Ramaskrishnan, V., Li, S., . . . Chen, C. (2012). miR-181d: a predictive glioblastoma biomarker that downregulates MGMT expression. *Neuro Oncol*, 712-719.
- Zhou, X., Du, Y., Jin, P., & Ma, F. (2015). Bioinformatic analysis of cancer-related microRNAs and their target genes. *Hereditas*, 855-864.
- Zhu, Y., Skogerbø, G., Ning, Q., Wang, Z., Li, B., Yang, S., . . . Li, Y. (2012). Evolutionary relationships between miRNA genes and their activity. *BMC Genomics*, 713-718.
- Zhuang, L., Xu, L., Wang, P., & Meng, Z. (2015). Serum miR-128-2 serves as a prognostic marker for patients with hepatocellular carcinoma. *PloS One*.
- Zisoulis, D., Lovci, M., Wilbert, M., Hutt, K., Tiffany, Y., & Pasquinelli, A. (2010). Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Mol Biol*, 173-179.

KASUTATUD VEEBIAADRESSID

- 1) mirBase – the microRNA database (kasutatud 04.05.17)
www.mirbase.org
- 2) HMDD v2.0: the Human microRNA Disease Database version 2.0 (kasutatud 28.06.17)
<http://www.cuilab.cn/hmdd>
- 3) USCS Genome browser (kasutatud 23.03.17)
<https://genome.ucsc.edu>
- 4) Human miRNA Expression database (kasutatud 03.07.17)
<http://guanlab.ccmb.med.umich.edu/mirmine/>

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Mirjam Tamme (sünnikuupäev: 04.07.1993)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Geneetiline varieeruvus mikroRNA geenides”, mille juhendaja on Tarmo Annilo (PhD)

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 21.08.2017